Superresolution of Magnetic Resonance Images

Nawar S. Alwesh
B.Sc. Eng.,
M.S.

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

2 October 2000

ABSTRACT

Magnetic resonance imaging (MRI) is used to image parts of the body using only electromagnetic interaction with the body's own atomic nuclei (hydrogen protons, in particular). The complex amplitude is directly measured at arbitrary locations in 2-D spatial frequency space; usually the locations are arranged in a square grid to allow reconstruction using the fast Fourier transform. The resolution of the obtained image is proportional to the number of locations and hence to the time allowed for data acquisition; and the bandlimit imposed by restricting the acquisition time is more severe in one of the coordinate directions than the other. To increase the efficiency of using expensive MR scanners and to reduce the time during which a patient must remain still, computational methods are sought to superresolve MR images.

Superresolution is commonly defined as the recovery of spatial frequency information of the object beyond the bandlimit imposed by the transfer function of the imaging system [Hunt 1995]. The well-known and straightforward Gerchberg-Papoulis (G-P) algorithm [Gerchberg 1974] was used to superresolve two sets of data, a simulated MR brain image and a phantom head image. These images were blurred by bandlimiting the spectrum (i.e. removing some of the high frequencies). Images resulting from applying the G-P algorithm were then compared to the corresponding original images. The algorithm was modified by making use of available information regarding the image, e.g. positivity, upper and lower bounds, and energy constraints, etc. The G-P algorithm and its modifications are special cases of the projection onto convex sets algorithm (POCS). The basic idea of POCS is that any prior information is used as a constraint on the image to lie in a closed convex set [Sezan and Stark 1982]. A more general form of the restoration algorithm, known as the method of generalized projections, extends the POCS method to utilizing nonconvex constraints, such as single level and neighbourhood based quantization. The different constraints were studied, and in some cases considerable improvement in the performance (as indicated by the improvement in the signal-to-noise ratio (ISNR)) and visually sharper images were achieved.

Since the superresolution problem is ill-posed [Hunt 1995] (i.e. trivial perturbation in the recorded data may lead to nontrivial perturbations in the solution), regularization methods are required to transform an ill-posed problem to a well-posed one, whose solution is an approximation to that of the ill-posed problem. In the case where no

information regarding the nature of the image was assumed, the well-known Tikhonov-Miller regularization method [Tikhonov and Arsenin 1977] was used to stabilize the problem by using a penalty function that represents a bound on the energy of the image. Other regularization methods such as constraining smoothness or limiting total variation (which either smooth out the solution or preserve edges) were studied. The choice of applying these methods depends on prior knowledge of the nature of the object.

The second part of the research concentrated on recovery of undersampled images, where the noisy spectra were undersampled by multiplying them with different undersampling patterns. The G-P algorithm, when applied on the undersampled images, in some cases resulted in useful recovery of even quite severely aliased regions. Both the determinacy of the system and the shape of the undersampling pattern used affected the amount of recovery achievable.

A study was also made on the recovery of a region of interest, which may be applied in situations where a follow-up image or dynamic imaging is required. The iterative region of interest (iROI) algorithm was presented in two versions. In the first version, information from a high resolution reference image is used to superresolve a low resolution dynamic image. This required some prior approximate knowledge of the location and extent of the region of interest. The total variation method was also used to improve the edges of the region of interest whenever accurate information regarding the region of interest was unavailable. The second version of the iROI method superresolves a low resolution dynamic image using a low resolution reference image; the resultant image is not as good as for the other version but it can be achieved with less prior knowledge.

Two other superresolving methods that utilize prior knowledge were investigated. The boxcar estimation method models the image by a series of boxcar functions with varying widths, locations and amplitudes. It results in sharp images with reduced Gibb's oscillation when applied to piecewise images. The nonconvex level penalty function utilizes prior knowledge of the levels present in the solution to superresolve an image. Both methods are shown to be useful when the image genuinely contains piecewise homogeneous regions.

ACKNOWLEDGEMENT

I would like to thank Dr. Philip Bones for his supervision, unlimited help throughout the course of this work, and for teaching me to always question my results. I would also like to thank John Connolly for his unlimited source of ideas, his cooperation, and for always helping me whenever I got stuck. I would also like to thank the University of Canterbury for its financial support through the University of Canterbury Doctoral Scholarship.

بسم الله الرحمن الرحيم

أود أن أقدم أطروحتي هذه إلى:
أبي وأمي الذين بإسنادهما وحبهما وتشجيعهما جعلا ذلك ممكنا. شكرا على كل ما منحتموني إياه، وسأبقى للابد مدينة لكما. أمل بعملي هذا أن يكون قد أثمر كل جهد بذلتموه وأنكما فخوران بي.

إلى بشار، زوجي العزيز المحب، الذي وضع ثقته بي وساندني في كل وقت عصيب. شكرا على تحملك خلال تقلبات مزاجي وعلى توفير كل احتياجاتي وعلى حبك الغير مشروط. لكن أهم كل شيء، شكرا على وضعك البسمة على وجهي. لم يكن هذا ممكنا بدونك. شكرا لعائلة الخطيب التي منحتني اكبر هدية على الإطلاق بشار، وعلى محبتهم ومعاملتهم لي كابنة.

إلي شقيقتي بسمة التي بدأت معها الحياة في عراك وكبرت لأحترمها والتمس المشورة منها. على الرغم من انك اصغر مني عمرا، لكنك ستكونين دوما مرساتي في الحياة. إلى عدي، شقيقي الصغير، الذي يبرر دوما مشاكله النفسية بكوني أخته الأكبر. سوف أرى دوما من خلال تعليقاتك الشقية، حبك واحترامك.

إلى دينا ودنيا، مرشداتي في الحياة، اللتان ساعدتا في نضوج الجانبين العملي والعاطفي في شخصيتي وتكاملها. إلى أختي الصغرى لينا، التي تعطي الكثير ولا تتوقع إلا القليل في المقابل. لقد كبرتي لتكوني المرأة الجميلة الحنونة التي توقعتها دوما. وشكرا لعائلة الحمداني على وجودهم لي خلال حياتي.

احبكم جميعا.

كذلك احب أن اقدم شكري لعائلة الصفواني، التي أصبحت عائلتنا الثانية في الغربة وعلى تجهيزهم الغير محدود للبسكويت ولكمال الصباغ على مساعدته وصداقته الحقيقية.

LIST OF PUBLICATIONS

During the course of this research, the following papers have been presented (and in some cases published in a refereed conference proceedings):

Bones, P.J., Alwesh, N. and Connolly, T.J. (1998), 'Superresolution of MR images', *Proc. Summer Topical Meeting, Signal Recovery and Synthesis, Optical Society of America*, Hawaii, pp.120-122.

Alwesh, N., Bones, P.J. and Connolly, T.J. (1998), 'Superresolution of MR images', *Proc. Image and Vision Computing New Zealand, IVCNZ98*, November, pp.246-251.

Alwesh, N. Connolly, T.J. and Bones, P.J. (1999), 'Recovery of undersampled magnetic resonance images', *Proc. Image and Vision Computing New Zealand, IVCNZ99*, August, pp.265-270.

Connolly, T.J., Alwesh, N. and Bones, P.J. (1999), 'Total variation based superresolution', *Proc. Image and Vision Computing New Zealand, IVCNZ99*, August, pp.283-288.

Connolly, T.J., Alwesh, A., and Bones, P.J. (2000), 'Utilizing prior knowledge for superresolution', *Australia and N.Z. Industrial Applied Mathematics Conference, ANZIAM 2000*, Copthorne Resort, Waitangi, Bay of Islands, NZ.

A paper on the original work presented in chapters 6 to 8 is currently under preparation for submission to an international journal.

CONTENTS

Chapter 1

INTRODUCTION

Magnetic resonance imaging (MRI) is based on the physical principles of nuclear magnetic resonance (NMR), which describes the interaction of those nuclei which have a nonzero magnetic moment, with a magnetic field. In 1920, W. Pauli postulated the concept that magnetic nuclei possess an angular momentum and magnetic moment, but it was not until 1946 that Felix Bloch and Edward Purcell independently, were able to measure NMR absorption in bulk materials. They were jointly awarded the Nobel prize in 1952. Between 1950 and 1970, NMR was only used for chemical and physical molecular analysis. After measuring NMR signals from live animals in 1967, J. Johns proposed using it on humans. This was only made possible in 1975 by P. Lauterbur who came up with the idea of using linearly varying fields to obtain spatially encoded signals. In 1976, Sir Peter Mansfield reported the first live human images and in 1981 the world's first commercial whole body MRI scanner was introduced. From 1980 to the present day, efforts have been directed towards improving both the software and hardware involved in the MR imaging process to obtain high contrast images with a high resolution in a very short time.

MRI is a non-invasive imaging technique that produces images of the internal physical and chemical characteristics of an object from externally measuring NMR signals [Liang and Lauterbur 1999]. It was initially used to image the brain and the spinal cord, but now is used to image joints (e.g. shoulder, wrist, ankle), blood vessels (e.g. renal arteries, leg arteries), breasts, internal organs (e.g. liver) and male and female reproductive systems. Unlike X-ray and computerized tomography (CT), MRI requires no harmful ionizing radiation and is capable of imaging along any orientation. When first used clinically a major disadvantage of MRI was that an imaging session used to take up to a few hours, but continued development has meant that imaging time has been reduced to as little as 10 minutes using a modern MRI 'scanner'.

In MR imaging, data is collected in the 2-D spatial frequency domain, which is then inverse Fourier transformed to produce the image. Since fine detail information is represented by the high spatial frequency portion of the spectrum, the resolution of the reconstructed image therefore depends on the amount of collected data which in turn

depends on the allowed imaging time (i.e. an increase in imaging time allows sampling in the spatial frequency domain to extend further, resulting in a more highly resolved image). In an effort to improve scanning economics (efficient use of the scanner) and patient comfort (e.g. patient is claustrophobic or a child), research has been directed towards reducing imaging time. Thus arises a need to superresolve images (obtained from reduced scan time) to obtain meaningful images for diagnostic purposes. The aim of the work presented in this thesis is to superresolve MR images (i.e. to recover spatial frequency information beyond bandwidth limit imposed by the reduced imaging time), with some emphasis on brain images. Different existing and new superresolving methods are applied to MR images in order to study their performance. The different methods are applied to both phantom and MR simulated brain images.

Chapter 2, of the thesis introduces the basic physical principles behind MR imaging followed by a description of the basic imaging sequence. Several fast imaging techniques are mentioned followed by description of the different artifacts resulting from either system imperfections or physical limitations.

Chapter 3 introduces the theory of superresolution and a proof of the possibility of achieving it. There are many different existing superresolving techniques. A very common approach to superresolution is to use prior information regarding the imaged object into the superresolution process. An example is the well-known and simply implemented iterative Gerchberg-Papoulis (G-P) algorithm, which utilizes prior knowledge of the support region of the object. Both the steepest descent and conjugate gradient methods are described and are related to the G-P algorithm. Results obtained from applying the G-P algorithm on both the phantom and MR image are presented.

Chapter 4 introduces the projection onto convex sets (POCS) method which uses additional prior information such as positivity, energy constraints, upper and lower bound constraint, etc. Different constraints may be used together in different combinations to superresolve an image. A more general form of POCS, known as the method of generalized projections (MGP) which utilizes nonconvex constraints, is discussed. New nonconvex constraints (single level and neighbourhood based quantization) are introduced and results obtained from applying these constraints are presented.

Since superresolution is an ill-posed problem by its nature, regularization methods are required to prevent iterative methods diverging. Chapter 5 discusses different regularization methods that may be used to transform an ill-posed problem into a well-posed problem that converges to a useful solution. Prior information regarding the object can be used in the regularization method to either smooth out the solution (e.g. reduce Gibb's oscillation and noise), preserve edges (using the total variation method) or a combination of both.

Up to this point in the thesis, superresolution of bandlimited images has alone been the subject of investigation. Chapter 6 studies the effect of undersampling the spatial

frequency domain using various undersampling patterns. Derivations of the aliasing patterns caused in the image domain are presented. The performance of applying superresolving methods to the various undersampled images is presented.

Chapter 7 discusses the use of superresolution algorithms for specific applications where only a region of interest (ROI) needs to be superresolved. Different existing algorithms (such as the keyhole and reduced field of view methods) are discussed and are compared to a new proposed algorithm, known as the iterative recovery of ROI method. The iROI method has two versions, in which either a high or low resolution image is used as a reference image. The total variation method is used to enhance the edges of the ROI.

Chapter 8 presents two further methods which utilize prior information in the recovery process. A parametric estimation method models the image by a series of boxcar functions of varying widths, amplitudes and locations, that need to be computed. A new method, known as the level penalty function, and based on the prior knowledge of the possible M levels that can be attained by the pixels, is introduced. Both methods are applied on both the phantom and brain data, and results are presented and discussed.

Chapter 9 presents a general summary of all the methods studied and conclusions based on all the results obtained. Suggestions for future research are also advanced.

Chapter 2

MAGNETIC RESONANCE IMAGING (MRI)

2.1 INTRODUCTION TO MRI

Magnetic resonance imaging (MRI) is a non-invasive imaging technique, used mainly in medical applications to produce high quality images of the inside of the human body. It is based on the nuclear magnetic resonance (NMR) principle which was first discovered by Bloch [Bloch 1946] and Purcell [Purcell *et al.* 1946] in 1945. They discovered that nuclei in chemically distinct environments resonate at slightly different frequencies in the presence of a magnetic field. MRI was once called nuclear magnetic resonance imaging, but the word 'nuclear' was dropped to avoid the misconception of the involvement of radioactivity. MRI scanners are used to image soft tissues of the body, especially the brain, spinal cord, nerves and muscles [Ballinger 1996]. The most important feature of MR imaging compared to other imaging techniques is the contrast between two soft tissues, which helps in distinguishing pathological tissues from healthy tissues. The first commercial MRI scanners were available around 1981. An MRI scanner requires no ionizing radiation and until now no adverse effects have been reported from its use.

2.2 IMAGING

The human body is made up mainly of water and fat, both of which contain high proportions of hydrogen atoms. The hydrogen nucleus has a single proton. This proton possesses a property known as the spin which can be modelled as a small magnet that will cause the hydrogen nucleus to exhibit NMR in the presence of a strong magnetic field. MRI is based on the interaction of the spin with three types of magnetic fields:

- a main magnetic field B_0;
- a radiofrequency (RF) field B_1; and
- linear gradient magnetic fields G.

By applying a uniform magnetic field $\mathbf{B}_0$(around 0.2 Tesla to 2 Tesla, for medical applications), oriented in space in the direction conventionally referred to as the z

direction, nuclei align themselves either parallel or anti-parallel to the field. Nuclei aligned parallel to the applied field are at a lower energy state than those aligned anti-parallel to the applied field. Transition between the two energy states occurs through the absorption of a photon with an energy E equal to the energy difference between the two states, where

$$E = \hbar\omega_0, \tag{2.1}$$

$\hbar$ is Plank's constant ($\hbar = 6.626\text{x}10^{-34} Js$) and ω_0 is the Larmor frequency, which is related to the applied field B_0 by

$$\omega_0 = \gamma B_0, \tag{2.2}$$

where γ is the gyromagnetic ratio, a constant for each type of atom. For hydrogen, γ = 42.58 MHz / Tesla.

There are fewer nuclei at the high energy state than those at the low energy state. Since the population difference is small and only the net magnetization of all nuclei contribute to any measurable signal, NMR signals are very weak. In a single volume element, which corresponds to a pixel in an MR image, there are many protons, each having a magnetic moment vector. The net magnetization $\mathbf{M} = M_x\mathbf{i} + M_y\mathbf{j} + M_z\mathbf{k}$ is the vector sum of the individual moments, where $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ are the unit vectors along x, y and z respectively, in the laboratory frame [Wright 1997]. In the absence of an external magnetic field, the spins are oriented randomly and the net magnetization $\mathbf{M} = 0$.

If an external magnetic field $\mathbf{B_0} = B_0\mathbf{k}$ is applied in the z longitudinal direction, the magnetic moment vectors will tend to align themselves either parallel or anti-parallel to the z direction (either at a low or high energy state) and will precess about the z-axis as shown in Fig. 2.1(a). The rotation, known as Larmor precession, occurs at a precise frequency known as the Larmor frequency given by Eqn. 2.2. Since the projection of the precessing moments on the xy transverse plane have a random phase, the net magnetization vector will therefore be along the z direction, with a vector length equal to the population difference of nuclei between the high and low energy states. However there is a slight excess of spins favouring the low energy state (i.e the parallel direction) therefore the net magnetization is $\mathbf{M} = M_0\mathbf{k}$, as shown in Fig. 2.1(b).

In the equilibrium state, the net magnetization is static and no current is induced in the receiver coil. To obtain information, the spins need to be excited. The excitation is achieved by applying a uniform RF pulse $\mathbf{B_1}$, orthogonal to the main magnetic field, in the xy (transverse) plane tuned to the resonant frequency of the spins to excite them out of equilibrium. This results in a deflection of the net magnetization away from the main field direction. The effect of applying $\mathbf{B}_1$ field is that the magnetization precesses about the $\mathbf{B}_1$ field away from the B_0 field direction as shown in Fig. 2.2(a). Note $\mathbf{B}_1$ ($\approx 50\mu Tesla$) is much weaker than $\mathbf{B}_0$ ($\approx 1.5 Tesla$) A rotating frame of reference, where x', y' and z are the axes of the rotating frame, that rotates about the

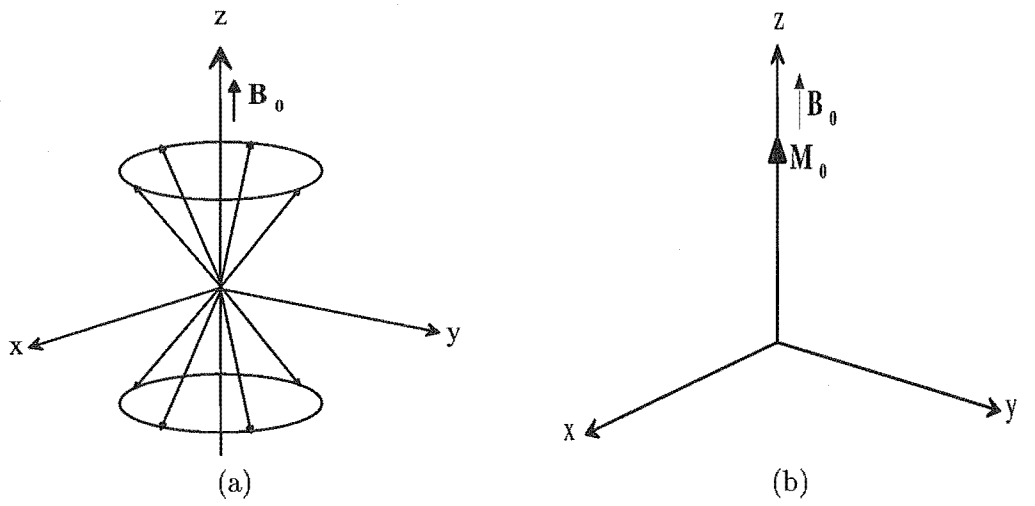


Figure 2.1 (a) Precession of the individual magnetic moments after applying $\mathbf{B_0}$; (b) net magnetization vector in the laboratory frame.

z axis at the Larmor frequency is used to simplify the visualization of the complex motions . Since $\mathbf{B_1}$ is tuned to the Larmor frequency, it therefore appears stationary in the rotating frame and is placed by convention along x'. Application of the RF pulse causes resonant absorption of energy, that excites some nuclei to the high energy state. The RF pulse $\mathbf{B_1}$ applies a torque on the magnetic moment vectors and tips them by an angle depending on the strength and duration of the RF pulse. The common excitation angle used is 90°, which tips the net magnetization vector to lie completely in the transverse plane, so that the longitudinal component $M_z = 0$ and the transverse component $M_{xy} \neq 0$ as shown in Fig. 2.2.

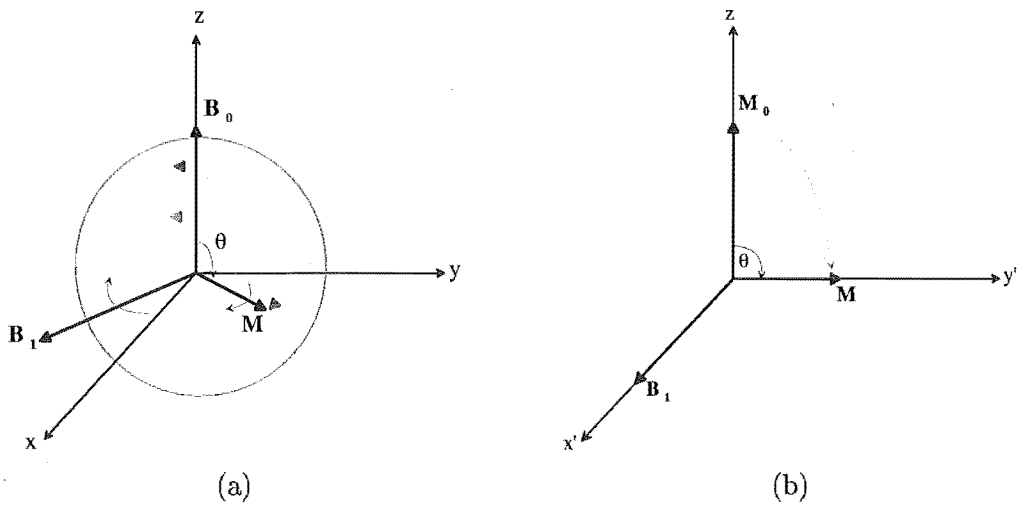


Figure 2.2 Precession of the magnetization after the application of both $\mathbf{B_0}$ and $\mathbf{B_1}$ in the: (a) laboratory frame; (b) rotating frame.

After the RF excitation, the nuclei return back to the equilibrium position by losing energy. This is known as the relaxation and starts at the termination of the RF pulse. The net magnetization M decays relatively slowly (in msec) to its equilibrium position along the longitudinal axis. This means that M_z increases from zero to an equilibrium value M_0 approximately according to

$$M_z = M_0(1 - e^{\frac{-t}{T_1}}), \tag{2.3}$$

where T_1 is the longitudinal relaxation time. This relaxation, known as the spin-lattice or T_1 relaxation, is due to the loss of excess energy resulting from the RF pulse to the surroundings as thermal energy. T_1 is a time constant which can be defined as the time required for the magnetization to return to 63% of its original equilibrium value. Typical T_1 values in biological tissues range from $300 - 2000msec$. The transverse component

$$\mathbf{M}_{xy} = M_x\mathbf{i} + M_y\mathbf{j} \tag{2.4}$$

has a magnitude M_{xy}, which decays to zero as equilibrium is achieved, that is approximately

$$M_{xy} = M_0 e^{\frac{-t}{T_2}}, \tag{2.5}$$

where T_2 is the transverse relaxation time. This relaxation is known as the spin-spin or T_2 relaxation, which ranges from $30 - 150msec$. This relaxation is due to the transfer of energy between nuclei in different energy states. The reason for this relaxation is that after applying the 90° RF pulse, all the nuclei are in phase, creating the transverse component of magnetization. Random local magnetic field variations due to interaction between the spins result in fluctuations in the precession frequency. This results in a random dephasing of the spins and the decay of M_{xy} by $e^{\frac{-t}{T_2}}$. Both relaxation processes occur simultaneously, with $T_2 \leq T_1$. A receiver coil specifically positioned to detect variations in the magnetization along the transverse plane detects a precessing magnetization and produces a signal known as the free induction decay (FID). A set of FIDs are collected and processed to construct an MR image.

2.2.1 The spin echo

In addition to the T_2 relaxation process, the transverse magnetization decays faster because nuclear magnetic moments do not all have the same precessing frequency due to imperfections of the applied magnetic field (ideally homogeneous) and added field gradients to achieve spatial resolution. This spread of resonance frequency leads to dephasing of the spins and a more rapid decay of the transverse magnetization than would occur purely by the T_2 relaxation phenomenon described above [Atlas 1996]. The exponential decay resulting from both the spin-spin (T_2) relaxation and the field inhomogenities is referred to as the effective relaxation (or T_2^*) time as shown in Fig.

2.3.

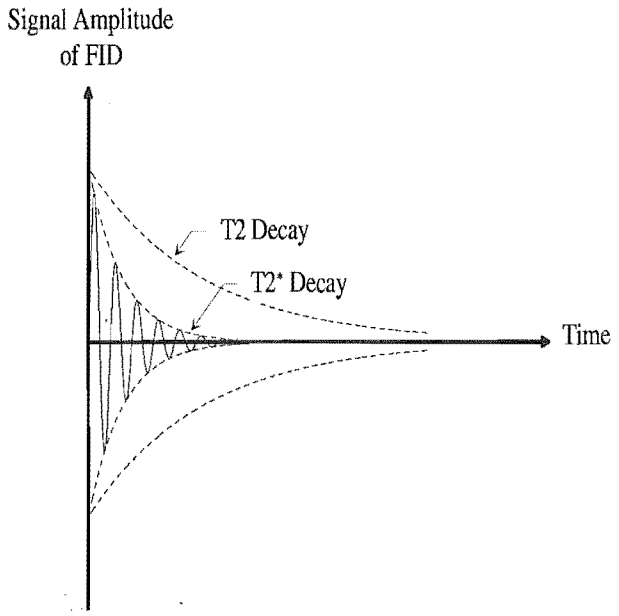


Figure 2.3 Effect of both T_2 and field inhomogenities on M_{xy} which is proportional to the FID amplitude.

The spin-spin interaction is random, time variant and irreversible. However, magnetic field inhomogenities have a constant influence on the spin system and can therefore be accounted for. The spin-echo sequence is used to remove the dephasing caused by the field inhomogeneity [Morris 1986]. In this sequence an echo, a reappearance of the signal a finite time after the disappearance of the FID [PHILIPS], is measured rather than the FID signal. A 90° RF pulse is applied along the x' direction, which flips the magnetization into the $x'y'$ plane along the y' axis as shown in Fig. 2.4(a). The transverse magnetization starts to dephase as shown in Fig. 2.4(b). After a time interval of $\frac{TE}{2}$, where TE is the time between the RF excitation and the maximum amplitude of the first echo, a 180° RF pulse is applied which flips the magnetization by 180° about the x' axis as shown in Fig. 2.4(c). The magnetization continues precessing in the same direction, which leads to at least a partial rephasing at time TE and resulting in an echo as shown in Fig. 2.4(d).

2.2.2 Selective excitation

In the situation described to this point, all the spins are resonating at the same frequency and are tipped by the same angle so that the excitation is non-selective. To image a thin slice of width Δz centred at z_0, selective excitation is required [Nishimura 1994]. This is achieved by applying a gradient field G_z, which is a variation in the magnetic

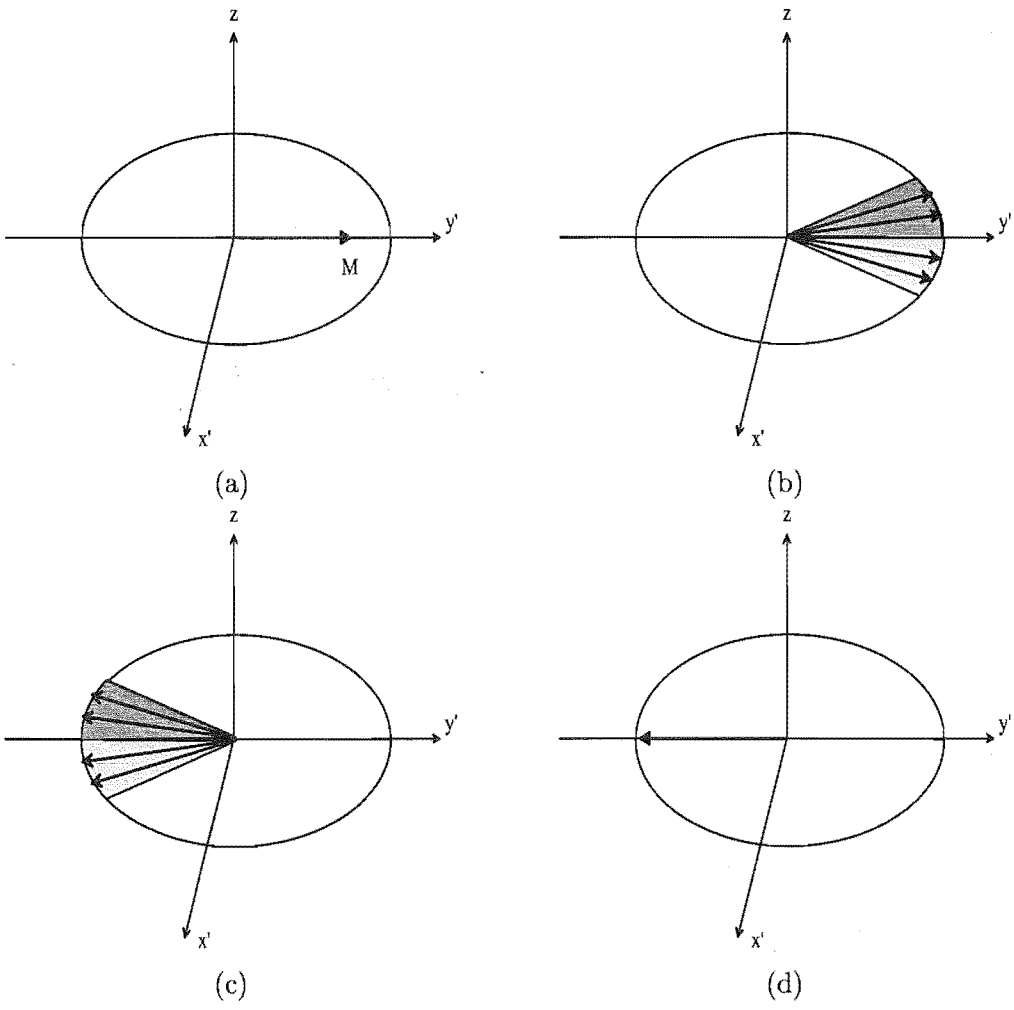


Figure 2.4 Spin-echo: (a) net magnetization after applying a 90o RF pulse; (b) dephasing of the spins; (c) flip of spins about x' axis after applying a 180° RF pulse at time $\frac{TE}{2}$; (d) rephasing of the magnetization at time TE.

field with respect to position, superimposed to $\mathbf{B}_0$ such that the net field

$$\mathbf{B} = (G_z(z - z_0) + B_0)\mathbf{k}. \tag{2.6}$$

This causes the resonant frequency of the spins to be a function of their z location. $B_1(t)$ must now possess a temporal frequency bandwidth that matches the bandwidth of resonant frequencies of spins in the slice of interest, to achieve a slice excitation with a rectangular profile as shown in Fig. 2.5. Thus the Fourier transform of $B_1(t)$ must be a rectangular function; hence ideally $B_1(t)$ must be a *sinc* function. In practice a compact approximation to a *sinc* function is used [Nishimura 1994].

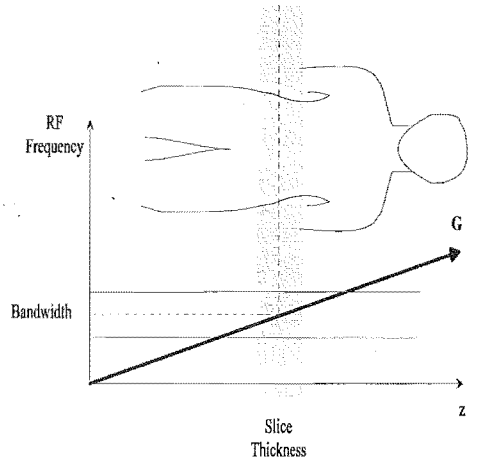


Figure 2.5 Slice selective excitation.

The applied G_z gradient not only results in selective excitation, which is desired, but also results in the undesired dephasing of the spins. In this case the dephasing is a linear function of z and can be removed by applying a gradient that is opposite in polarity to the originally applied G_z field. This is known as refocusing or time reversal gradient.

Once a thin slice of interest has been isolated using selective excitation, gradient fields need to be applied to get spatially dependent information within the slice. A linear gradient field G_y is applied for a time τ_y such that $\mathbf{B} = (G_y(y) + B_0)\mathbf{k}$. This is known as the phase encoding pulse where $G_\phi(t) = G_y(t)y$. It causes the phases of the spins, within the selected slice, to be a function of their y location. Therefore the phase of a spin $\Phi(y)$ at the completion of the pulse is

$$\begin{aligned} \Phi(y) &= \gamma \int_0^{\tau_y} (B_0 + G_y(t)y)dt = w_0\tau_y + 2\pi k_y(\tau_y)y \\ \text{where} \quad k_y(\tau_y) &= \frac{\gamma}{2\pi} \int_0^{\tau_y} G_y(t)dt \end{aligned} \tag{2.7}$$

.

where k_x, k_y and k_z are the spatial frequency variables.

Another linear gradient field, known as the frequency encoding field, is applied such that $\mathbf{B} = (G_x(x) + B_0)\mathbf{k}$, where $G_f(t) = G_x(t)x$, which causes the frequency, $\omega = \omega(x,t)$,of the spins to be a function of their x location, where

$$\begin{aligned} \omega &= \gamma(B_0 + xG_x(t)) = \omega_0 + \gamma x G_x(t) \\ k_x(t) &= \frac{\gamma}{2\pi}\int_0^t G_x(t)d\tau. \end{aligned} \tag{2.8}$$

The signal received contains the contributions of the spins in the selected slice with the phases depending on the y location of the spins and frequencies depending on the x location of the spins.

The most common and current imaging technique used is the spin-warp sequence version of the 2-D Fourier transform (FT) imaging [Felix *et al.* 1988]. Fig. 2.6 shows a simple timing diagram for the spin-warp sequence, which shows the 90° RF pulse, the slice selection gradient pulse G_z, the phase encoding pulse G_ϕ, the frequency encoding pulse G_f, and the received signal. The first stage of the imaging process is applying the 90° RF pulse and the G_z pulse at the same time. G_z is turned off when the RF pulse is completed. All the nuclei within the selected slice are excited. Unwanted dephasing occurring during the slice selection is removed by applying a time reversal gradient(explained previously). The phase encoding pulse is then turned on. The amplitude of the phase encoding gradient is different for each readout. A 180° RF pulse is then applied to refocus the dephasing caused by field inhomogeneity as discussed above. The spin echo signal is then collected. It is important to note that since both lobes of the readout gradient are on opposite sides of the 180° RF pulse, they both have the same polarity. For the next readout, G_y or τ_y is varied and the same gradient G_f is used. Any given G_y or τ_y value corresponds to a value of k_y (a horizontal line in k-space as shown in Fig. 2.7(c)). The value of k_x varies from one limit to another as a function of the time during the readout process. The procedure is repeated several times, each time varying G_y or τ_y in equal steps, until enough information in k-space is collected. The corresponding k-space trajectory of the above sequence is shown in Fig. 2.7(a). For a specific value of G_y, the k-space trajectory moves along the k_y axis (position 1). When a positive G_x gradient is applied, the k-space trajectory then moves to the right along the k_x axis (position 2), resulting in some spatial frequency (position 3). Applying the 180° pulse flips the k-space position to its conjugate position as shown in Fig. 2.7(b). The sweeping along the k_x direction begins when the readout pulse is applied. Once all the information in the k-space is collected (Fig. 2.7(c)), a 2-D Fourier transform is performed on the data collected in the k-space to obtain the MR image [Hornak 1996] shown in Fig. 2.7(d).

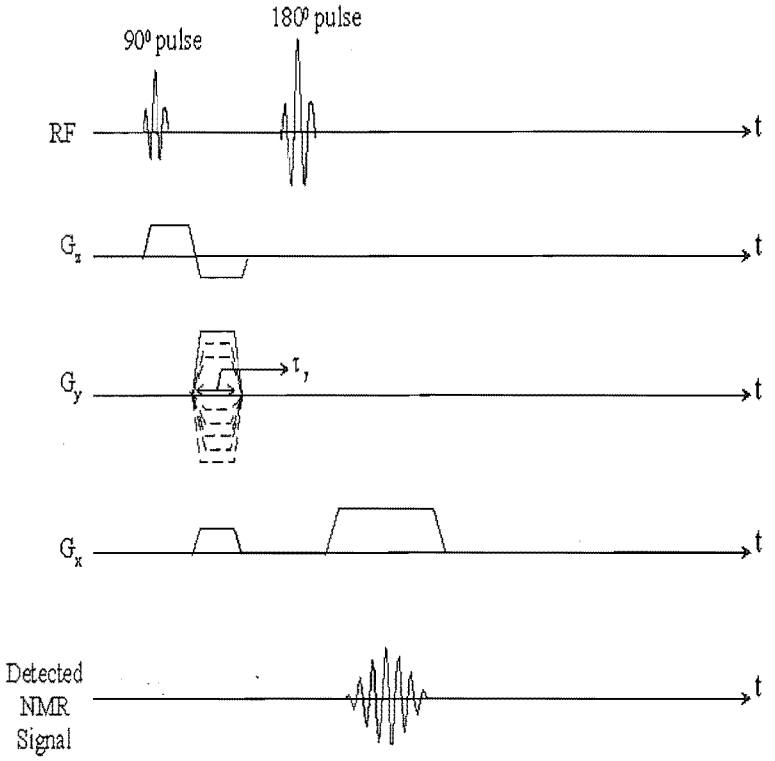


Figure 2.6 Timing diagram for a simple FT imaging technique.

2.2.3 Inversion Recovery (IR) Imaging Method

One of the important attributes of MR imaging is the ability to achieve a strong contrast between tissues of interest by adjusting the imaging parameters. Spin density, T_1 and T_2 all vary between different tissues and affect signal strength. By changing imaging sequences, the differences between the parameters can be exaggerated. For example, if the imaging parameters are adjusted such that T_1 has the dominant effect, then this is referred to as 'T_1 weighting'.

The inversion recovery imaging method is a well known T_1 weighted imaging sequence as shown in Fig. 2.8. In this case a 180° pulse is applied which inverts the magnetization from along the positive z direction to the negative z direction. After applying the pulse, the spins are allowed to relax for a time TI and the longitudinal magnetization M_z starts to increase form the most negative value to zero and builds up in the positive direction. At time TI, the magnetization is [Nishimura 1994]

$$M_z = M_0(1 - 2e^{-\frac{TI}{T_1}}), \tag{2.9}$$

where M_0 is the equilibrium value. At time TI, a 90° excitation pulse is applied which flips M_z into the transverse plane, therefore $M_z = 0$. At time TR, a 180° pulse is then applied to invert the magnetization. Between the times TI and TR (before the 180° pulse), M_z builds up from zero to $M_0(1 - e^{-\frac{TR-TI}{T_1}})$ and at time TR (after the 180°

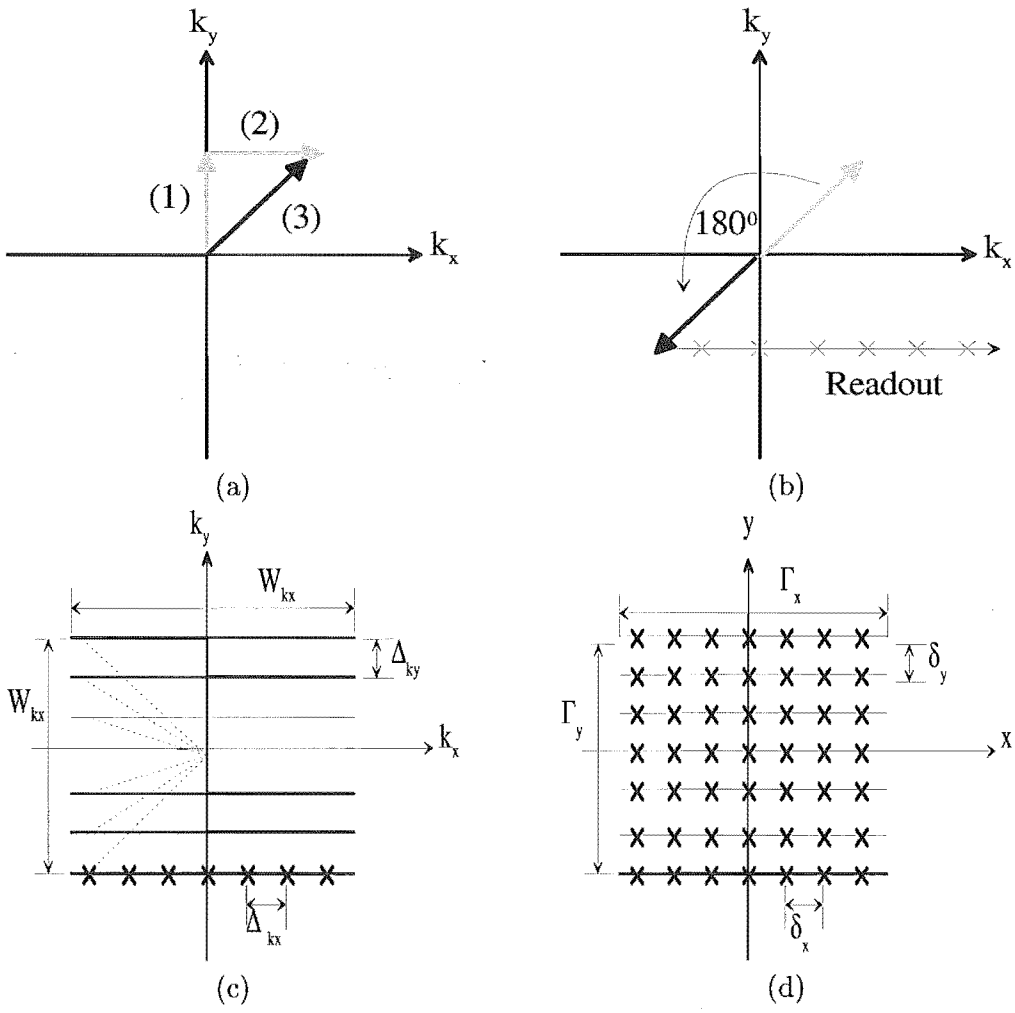


Figure 2.7 (a) Effect of applying the gradient pulses; (b) flip in the k-space due to the 180° pulse; (c) collected data in the k-space; (d) image.

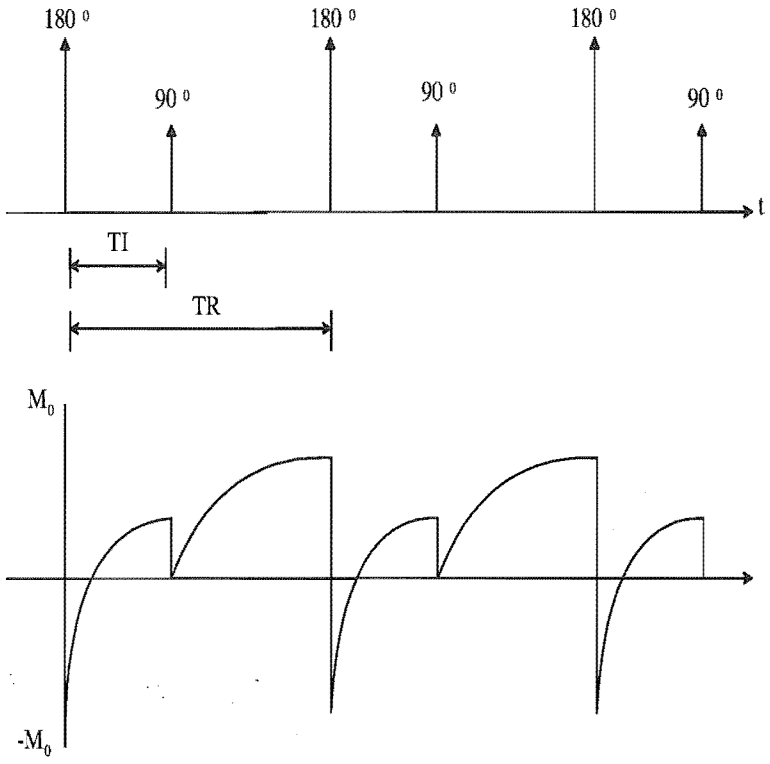


Figure 2.8 Timing diagram for the inversion recovery imaging technique.

pulse)

$$M_z = -M_0(1 - e^{-\frac{TR-TI}{T_1}}). \tag{2.10}$$

M_z starts to recover again according to

$$M_z = M_0(1 - 2e^{-\frac{TI}{T_1}} + e^{-\frac{TR}{T_1}}). \tag{2.11}$$

In this case, TI and TR can be adjusted so that specific tissues will have zero or negative intensities. For example, if $TI = (\ln 2 - \ln(1 + e^{-\frac{TR}{\tau}}))\tau$, then any tissues with $T_1 = \tau$ will have a zero intensity [Hoult and Lauterbur 1979].

2.3 FAST IMAGING TECHNIQUES

When the idea of using MR imaging in clinical applications was first introduced, there was a lot of criticism due to it being a slow technique (imaging time lasted up to tens of minutes). However, since then, major developments have been made by using different techniques to reduce scan time. Initially, the two main reasons for wanting to reduce scan times were efficient use of the expensive scanner and patient comfort. However with the introduction of fast imaging techniques, another advantage was found. With imaging time reduced to fractions of a second, physiological motion could be virtually

frozen and studied. In addition physiological motion (e.g. respiration) no longer degraded the image. In the conventional method (as described in section 2.2.2), scan time T_S is the product of the time interval between the RF excitation pulses T_R and the number of phase encoding steps N_P. Therefore, one of the approaches used to reduce scan time is reducing the number of samples in the phase encoding direction, causing the collected data to be asymmetric with usually 256 samples in the frequency encoding direction and 128 samples in the phase encoding direction (resulting in a corresponding loss in resolution). Another approach to reduce scan time is reducing T_R. This approach is used in the fast low angle shot (FLASH) imaging technique explained below. The fast imaging techniques can be divided into two types:

- Single shot imaging: a complete set of data is collected by applying a single RF excitation pulse followed by a series of gradient pulses. This imaging technique can be further subdivided into echo planar imaging (EPI) and spiral scan echo planar imaging (SEPI).
- Multiple shots imaging: T_R is reduced by applying a series of RF excitation pulses during which a single data acquisition sequence is applied. An example of this imaging technique is fast low angle shot (FLASH) imaging.

Each of the above mentioned techniques is described below.

Section 2.3.4 describes a method that can be used in both the conventional and fast imaging techniques to reduce imaging time by utilizing the symmetry of the raw data.

2.3.1 Echo planar imaging (EPI)

In the conventional imaging technique, a single line in the k-space is scanned after applying a single RF excitation pulse. The EPI technique aims at collecting a complete set of data by applying a single RF excitation pulse followed by specifically designed gradient pulses, as shown in Fig. 2.10(a). In the EPI method, a series of short duration phase gradient pulses known as 'blips' are applied. The trajectory in the k-space for the corresponding sequence is shown in Fig. 2.10(b). The trajectory in the k-space starts at time T_A, where $k_x = 0$ and $k_y = 0$. Between time T_A and time T_B, both a phase encoding and a frequency encoding pulse are applied, setting $k_x = k_{xmax}$ and $k_y = k_{ymax}$. The k-space trajectory is flipped by applying a 180° pulse, setting $k_x = -k_{xmax}$ and $k_y = -k_{ymax}$. A readout gradient is then applied between times T_C and T_D, which scans along $k_y = -k_{ymax}$. The application of the blip at time T_D, increments k_y by Δ_{k_y}. Scanning along $k_y = -k_{ymax} + \Delta_{k_y}$ is achieved by applying a second readout gradient between times T_D and T_E [Atlas 1996]. One disadvantage of the EPI method is the complex hardware required to handle high sampling rates (since a complete set of data is collected during a single excitation) and to generate

fast switching gradients. The EPI method is utilized in functional MRI, the imaging of the cerebral physiology. As mentioned previously for conventional imaging methods, the EPI has a lower resolution in the k_y direction since only a limited number of phase encodes can be generated.

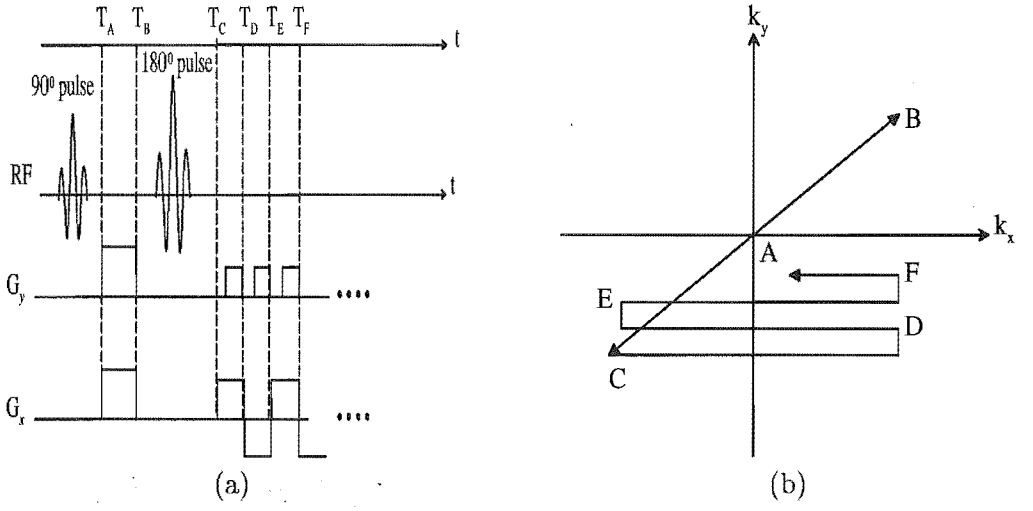


Figure 2.9 EPI method: (a) pulse sequence; (b) k-space trajectory.

2.3.2 Spiral scan echo planar imaging (SEPI)

To overcome the poor resolution problem in the k_y direction of the EPI method, the SEPI method aims to scan the k-space in a spiral trajectory to obtain a circularly symmetric k-space scan. To obtain a regular spiral in k-space, sinusoidal signals can be utilized as follows [Cho *et al.* 1993, Ahn *et al.* 1986]

$$\begin{aligned} k_x(t) &= \gamma\eta t cos(\varsigma t) \\ k_y(t) &= \gamma\eta t sin(\varsigma t), \end{aligned} \tag{2.12}$$

where η and ς are constants to be determined. Therefore

$$\begin{aligned} G_x(t) &= \frac{1}{\gamma}\frac{d}{dt}k_x(t) = \eta cos(\varsigma t) - \eta\varsigma t sin(\varsigma t) \\ G_y(t) &= \frac{1}{\gamma}\frac{d}{dt}k_y(t) = \eta sin(\varsigma t) + \eta\varsigma t cos(\varsigma t). \end{aligned} \tag{2.13}$$

The k-space can be represented in the polar coordinate system $(k_r(t), k_\theta(t))$ as shown in Fig. 2.10 by using the following transformation:

$$\begin{aligned} k_r(t) &= \sqrt{k_x(t)^2 + k_y(t)^2} = \gamma\eta t \\ k_\theta(t) &= tan^{-1}\frac{k_y(t)}{k_x(t)} = \varsigma t. \end{aligned} \tag{2.14}$$

Let Δ_t be the time required to increase $k_\theta(t)$ by $\Delta_{k_\theta}(t)$ and N_θ be the number of

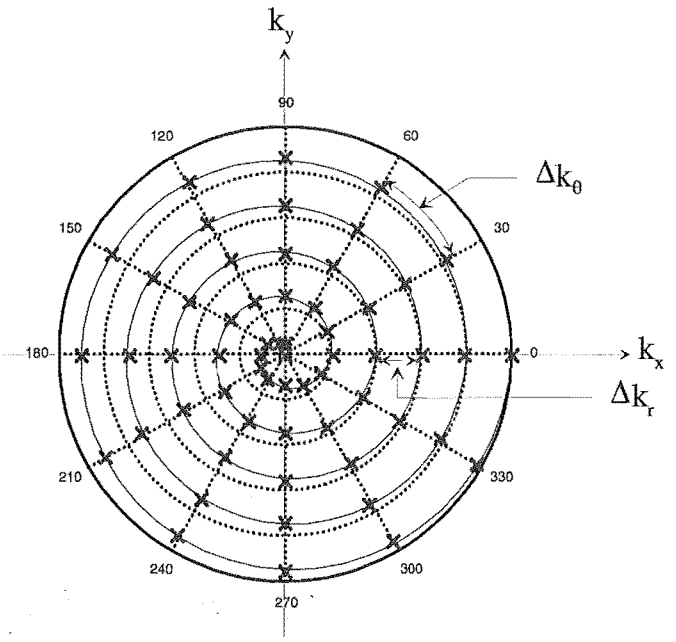


Figure 2.10 k-space trajectory in SEPI method pulse.

samples in one complete rotation. The radial and angular increments $\Delta_{k_r}(t)$ and $\Delta_{k_\theta}(t)$ can be defined as

$$\begin{aligned} \Delta_{k_r}(t) &= \gamma\eta N_\theta \Delta_t = \frac{2\pi}{2N_r\Delta_r} \\ \Delta_{k_\theta}(t) &= \varsigma\Delta_t = \frac{2\pi}{N_\theta}, \end{aligned} \tag{2.15}$$

where N_r is the number of rotations in the k-space and Δ_r is the spatial resolution. From Eqn 2.15, the constants η and ς can be determined as

$$\begin{aligned} \eta &= \frac{\pi}{\gamma N_\theta N_r \Delta_t \delta_r} \\ \varsigma &= \frac{2\pi}{N_\theta \delta_t}. \end{aligned} \tag{2.16}$$

The maximum radial frequency k_{rmax} can be defined as

$$\begin{aligned} k_{rmax} &= N_r\Delta_{k_r} = \gamma\eta T_S \\ \text{where} \quad T_S &= N_r N_\theta \Delta_t \end{aligned} \tag{2.17}$$

where T_S is the total scan time. Since spatial resolution depends on k_{rmax}, it therefore depends on the number of rotations in the SEPI method. To utilize the fast Fourier transform, the SEPI data in the polar coordinates needs to be interpolated and resampled onto a Cartesian coordinate system. Many different forms of spiral imaging can

be found in various literature [Duyn and Yang 1997, Bornert *et al.* 1999].

2.3.3 Fast low angle shot (FLASH) imaging

This imaging technique is very similar to that of the conventional imaging method. The major difference between the two techniques is the use of an $\theta°$ RF excitation instead of the 90° RF excitation used in conventional imaging methods. In conventional imaging, if a second 90° RF excitation pulse is applied at a time less than T_1, the signal produced by the second excitation will be less than that produced by the first by $e^{\frac{-T_R}{T_1}}$. If a series of 90° RF pulses are applied, the signal in steady state will be

$$M_z = M_0(1 - e^{\frac{-T_R}{T_1}}). \tag{2.18}$$

This is known as saturation and will lead to signal suppression in conventional imaging. Since scan time T_S depends on T_R, a reduction in T_R causes a reduction in scan time. In the FLASH imaging method, the angle $\theta°$ is related to $B_1(t)$ as follows:

$$\theta = \gamma \int B_1(t)dt. \tag{2.19}$$

If a set of $\theta°$ RF pulses are used, the transverse magnetization M_{xy} is defined as [Atlas 1996]

$$M_{xy} = \frac{1 - e^{\frac{-T_R}{T_1}}}{1 - cos(\theta)e^{\frac{-T_R}{T_1}}}. \tag{2.20}$$

From Eqn 2.20 it is obvious that by reducing θ, the signal becomes less dependent on T_1. Therefore T_R can be reduced without causing saturation and signal suppression. The FLASH imaging technique is used in imaging dynamic effects.

2.3.4 Half Fourier imaging technique

Another technique used in reducing imaging time is based on the symmetry of the raw data. In this case only one half of the k-space is sampled by measuring $\frac{N}{2}$ samples in the phase encoding direction and N samples in the frequency encoding direction. Since the k-space has conjugate symmetry i.e. $k_{ij} = k^*_{-i-j}$ [Feinberg *et al.* 1986], the values of the remaining half of the k-space are filled before performing the Fourier transform to reconstruct an $N \times N$ image. This technique may be applied to any imaging sequence such as the conventional or fast imaging techniques.

Ideally, the conjugate symmetry property of the raw data is valid resulting in a real image. However, in practice magnetic field inhomogeneities, object motion, fat-water chemical shift and magnetic susceptibility (discussed in section 2.5.1), introduce

frequency shifts in the k-space domain which correspond to spatially dependent phase variations in the image domain resulting in complex images. The phase correction required for the image obtained through the half Fourier imaging is the same correction required for the normal image (obtained through symmetrical k-space acquisition) and since phase varies slowly across a normal image, only a low resolution image is required. In order to estimate the image phase, instead of measuring only half of the k-space, a few extra lines are measured in the other half of the k-space, e.g the rows $k_y = -n_0, \ldots, \frac{N}{2} - 1$ are measured, where $n_0 << \frac{N}{2}$. The central data, $-n_0 \leq k_y \leq n_0$ is then zero padded to obtain a $N \times N$ matrix. To reduce Gibbs oscillations, a Hamming window is used to soften the sharp transitions between the measured data and the zeros. A low resolution image is then obtained by inverse Fourier transforming the modified data. The phase corrections needed to make the low resolution image real are then used to make the image resulting from the half Fourier measured data real [Margosian *et al.* 1986]. It is important to note that applying the Hamming window lowers the image amplitude and that the signal to noise ratio (SNR) of the $\frac{N}{2} \times N$ image resulting from the half Fourier imaging technique is decreased by a factor of $\sqrt{2}$ compared to the $N \times N$ image resulting from the symmetrical k-space sampling [Hoult and Lauterbur 1979]. In addition, the quality of the reconstructed image from the half Fourier imaging method depends on how well the phase of the low resolution image estimates the phase of the image that would have been obtained from an $N \times N$ symmetric k-space data.

2.4 HARDWARE

The MRI system is divided into four main parts:

- magnet;
- gradient coils;
- RF coils and transceiver; and
- data processing system.

Each of these is now described briefly.

2.4.1 Magnet

A MRI scanner is based on a large electromagnet, which produces the main static magnetic field described in the previous section. The magnet is the main factor in determining the cost, appearance and capabilities of the MRI system. The main characteristic

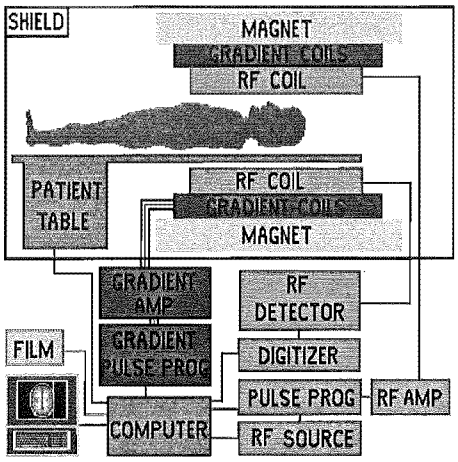


Figure 2.11 A simplified schematic diagram of a MRI scanner.

of the magnet is the strength of the magnetic field produced within the patient. Also important are the uniformity and temporal stability of the magnetic field.

A magnetic field is generated a magnetic field, either by using a ferromagnetic material or by passing an electrical current through a wire. The former method is more efficient and produces less stray fields, but it has the disadvantage of having a limited design flexibility and is very heavy. Most MRI scanners use the electrical current as a source of magnetic field. The simplest design is the solenoid, which produces large volumes of uniform field and is very flexible. The resistive or superconductors are the two types of conductors used. The resistive magnet is simple and has a low cost to construct, however it only produces small regions of low field strength when considerable power is used. Superconducting magnets are more difficult and expensive to construct, but they produce large volumes and high fields [Felix *et al.* 1988]. Summarizing the above, there are four different types of magnets available for MRI.

- Permanent magnets, which do not require an external power source, are made from 'hard' magnetic materials, which are difficult to magnetize, but once magnetized maintain their magnetization for a long period.

- Electromagnets are made from 'soft' magnetic materials, which are easily magnetized by applying a current, but lose their magnetization when the current supply is terminated.

- Resistive magnets, which require current flow in metallic wires to generate a magnetic field without the use of any magnetic materials. They are inexpensive and are simple to construct, but generate heat due to the currents in the windings.

- Superconducting magnets require no external power source once energised, and have no resistance when cooled to below their transition temperature. These

magnets are currently widely used in MR scanners producing fields between 0.15 and 4 Teslas.

All of the above magnets and hybrid magnets have been successfully used for MRI. The superconducting magnets have dominated MRI since the mid 1980s, due to their ability to reach high field strengths and achieve a high degree of field homogeneity. Superconducting magnets need cooling to reach their critical or transition temperatures. This is achieved by using liquid Helium as the refrigerant, where considerable effort is made to minimize heat loss from the Helium vessel by reducing the number of conduction, convection and radiation paths [Webb 1995]. A typical superconducting magnet has about 17000 turns of wire with a radius of about 0.65m. The wire and its supporting structures weigh about 3 tons.

2.4.2 Gradient coils

Within the magnet are three gradient coils, each with a separate power supply and independent computer control, which produce the gradient fields allowing slice selection and encoding of spatially-dependent information into the MR signal. Many coil shapes can be used, but in MR scanners using superconducting magnets, it is common to wind all three coils into a cylinder surrounding the patient. The main properties of the gradient coils is to make the magnetic field a function of position. The effects of the gradient fields should be strong enough to overcome the effects due to inhomogenities of the main applied field. However, high gradient fields are difficult to produce and switch quickly, and need a bigger receiver bandwidth for data collection, which in turn lowers the signal to noise ratio of the received signal. Gradient coils must also produce fields with high linearity, which controls the accuracy of the collected data. Nonlinearity of the gradient fields results in spatial distortions of the image. Gradient fields must have a short rise time (about 1msec), which is achieved through a complex process since the gradient coils are placed in the middle of a magnet. Due to the switching of gradient fields, Eddy currents which oppose the fields produced by the gradient coils, are produced in surrounding materials, resulting in poor image quality. To reduce the effect of the Eddy currents, the gradient coils are driven by complex waveforms that depend on the type of magnet and imaging sequence being used.

2.4.3 RF coils and transceiver

Within the gradient coils is a RF coil that produces the RF field and detects the received signal at and near the Larmor frequency as shown in Fig. 2.11. Since the frequency range used in MR imaging is the same as that used in radio and television broadcasting, the room housing the scanner is surrounded by an RF shield that prevents the RF pulses radiating outside the room and prevents other RF signals radiating into

the scan room [Hornak 1996]. The RF transmitter generates a high burst of RF energy (for a few msecs), it then detects the received signal (which lasts from 10-1000msecs). The same coil can be used for both transmitting and receiving a signal, in which case careful protection of the preamplifier in the receiving system is required from the instantaneous burst of energy during transmission. Use of separate coils for the transmitting and receiving processes has the advantage of individually optimizing the coil design for each function, however care should be taken to prevent interaction since both coils operate at the same frequency

The purpose of the transceiver is to produce the appropriate RF voltage to the RF coil and to detect the FIDs. During the receive state, the RF coils are connected to a low noise preamplifier, and then to the main amplifier. Complex fast switching circuitry is required to protect the preamplifier from any signals from the transmission state from leaking and damaging it. The received signal is then digitized using analogue to digital (A/D) converters and then sent to the computer for processing.

2.4.4 Data processing system

The data system can be divided into two parts. The first is the computer, which is responsible for many tasks such as acquiring data, generating pulse sequences, hardware control, patient table control, safety monitoring, image display etc. The second part of the data system comprises the subsystems that, under the direction of the computer, perform specific functions. Some of the main subsystems are:

- The pulse programmer: is responsible for timing different events that make up the imaging sequence. The important properties of this subsystem are accuracy, speed, flexibility and ease of programming.
- The array processor: is responsible for generating the images by performing the Fourier transforms and all necessary tasks related to image processing.
- Image Display: is by means of cathode ray tubes using an image processor, typically with a display matrix of 512 x 512. The image processor enables choosing a region of interest by a cursor and performing various functions, such as addition, subtraction etc.

2.5 ARTIFACTS

MR imaging has developed rapidly in the last decade and has proven itself to be a very reliable diagnostic tool. It is comparatively new to other existing imaging tools such as ultrasound and X-ray imaging and has therefore introduced new artifacts that need to be investigated. Artifacts can be defined as any aspect of an image which misrepresents the anatomic and geometrical relationships within the body [Atlas 1996]. There are

two important purposes for investigating artifacts and their sources: 1) to avoid false diagnoses being made based on artifacts, and 2) in order to learn how to eliminate them. Some of the artifacts and their sources can be easily identified, while some are difficult to interpret. The main sources of artifacts are

- frequency shifts;
- truncation;
- sampling and aliasing;
- instrumental errors;
- motion;
- noise; and
- measurement errors.

Each of these is discussed below.

2.5.1 Frequency shifts

Ideally the resonant frequency is uniform across a sample, however in practice that is not the case due to several reasons:

- field inhomogeneity: since the Larmor frequency is a function of magnetic field strength, any nonuniformity in the main applied magnetic field will lead to a nonuniformity of the resonant frequency, which will cause distorted images. The distortion can be either spatial (i.e. shift in position of the object), intensity or both. This artifact can be reduced by increasing gradient strengths or by shimming the magnetic field.

- fat-water chemical shift: since both fat and water contain protons that are in different chemical environments, they will resonate at slightly different frequencies. Using frequency encoding on part of a body that has both fat and water at the same location will produce an image with the water and fat slightly shifted from each other. The displacement is dependent on the strength of the applied magnetic field. Therefore, high field systems will demonstrate the artifact more severely. The artifact will appear as a dark area where the physical signal present here is shifted, and a brighter region where both the signals due to the fat and water overlap. This artifact can be minimized by using chemical shift imaging, which images the water and fat components independently [Webb 1995].

- magnetic susceptibility: this term describes the phenomenon in which a material becomes magnetized when subjected to a magnetic field. The relationship between the magnetic field strength H and magnetization M is $M = \chi H$, where χ is the magnetic susceptibility. Even in the case of a perfectly homogeneous magnet, the resonant frequency will still be nonuniform across a sample due to the difference in susceptibility within the sample [Atlas 1996]. This distortion due to the difference in susceptibility of two adjacent materials is more obvious at their boundaries (e.g. interface between air and tissue). This artifact appears either as an increased signal intensity at the interfaces or as a reduced signal intensity within a voxel (e.g. lung imaging). This extent of this artifact can be diminished by increasing the gradients' strengths.

2.5.2 Truncation

Truncation artifacts appear as low intensity streaks usually close to moderately straight tissue interfaces. Due to the regularity of the artifact, they can be easily misinterpreted as nerve fibers, muscle bundles, etc. The truncation artifact is also known as Gibbs phenomenon, spectral leakage or edge ringing. Since the object being imaged has sharp discontinuities (including the interface with the surrounding air) its spatial frequency spectrum extends to infinity. The measurements in k-space are, however, necessarily restricted in extent. The resulting truncation of the collected data results in Gibbs phenomena in the reconstructed image. The truncation artifact is more obvious in the phase encoding direction where the truncation is more severe due to imaging time limitations. A direct solution to the artifact is to obtain more measurements of the high frequency information, but that requires a longer imaging time [Henkelman and J.Bronskill 1987]. Truncation artifacts can be reduced by imposing a filter (window function) on the measurements in the k-space to smoothly roll off the highest frequencies. This eliminates the discontinuity in the spectrum, which reduces the edge ringing, but also results in blurring of the image. The filter must be carefully designed to achieve a suitable compromise between the blurring and the ringing effect. Many procedures to extrapolate or recover the data beyond the cutoff frequency (to superresolve) are available. Such methods are the topic of chapters 3 to 8 in this thesis.

2.5.3 Sampling and aliasing

In digital image processing, an array of data is obtained by sampling a physical parameter (e.g. intensity, density, complex amplitude, etc.) representing the physical image. In MR imaging, sampling occurs in the spatial frequency domain (k-space) as explained earlier in this chapter. This can be modelled as multiplying the spectrum $F(k_x, k_y)$ by

a sampling function $S(k_x, k_y)$ consisting of a series of Dirac delta functions,

$$S(k_x, k_y) = \frac{1}{\Delta_{k_x}\Delta_{k_y}} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(k_x - m\Delta_{k_x}, k_y - n\Delta_{k_y}), \tag{2.21}$$

where Δ_{k_x} and Δ_{k_y} are the sampling intervals and $\frac{1}{\Delta_{k_x}}$ and $\frac{1}{\Delta_{k_y}}$ are the sampling rates along the k_x and k_y directions respectively. The sampled spectrum is thus

$$\hat{F}(k_x, k_y) = F(k_x, k_y)S(k_x, k_y). \tag{2.22}$$

Since multiplication in the frequency domain represents convolution in the space domain, the inverse Fourier transform of $\hat{F}(k_x, k_y)$ can be written as

$$\begin{aligned} \hat{f}(x, y) &= \mathcal{F}^{-1}(\hat{F}(k_x, k_y)), \\ &= \frac{1}{\Delta_{k_x}\Delta_{k_y}} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(x - \frac{m}{\Delta_{k_x}}, y - \frac{n}{\Delta_{k_y}}) \end{aligned} \tag{2.23}$$

where $\mathcal{F}^{-1}$ represents the inverse Fourier transform operator and $f(x, y)$ is the inverse Fourier transform of $F(k_x, k_y)$. It is obvious from Eqn 2.23 that the image $\hat{f}(x, y)$ is replicated infinitely with separation $\frac{1}{\Delta_{k_x}}$ and $\frac{1}{\Delta_{k_y}}$ in the x and y directions respectively. The spacing between the replications is known as the field of view Γ, which is defined as (see Fig. 2.7(c))

$$\begin{aligned} \Gamma_x &= \frac{1}{\Delta_{k_x}} \quad \text{field of view in x direction} \\ \Gamma_y &= \frac{1}{\Delta_{k_y}} \quad \text{field of view in y direction.} \end{aligned} \tag{2.24}$$

As explained in section 2.2.2, for the phase encoding process, the amplitude of the phase encoding gradient is G_y, which is modified by ΔG_y between successive rows in k-space. Therefore, the k-space sampling interval Δ_{k_y} is

$$\Delta_{k_y} = \frac{\gamma}{2\pi}\Delta_{G_y}\tau_y. \tag{2.25}$$

In the frequency encoding (x) direction, the incremental gradient area Δ_{k_x} depends on the amplitude of the frequency encoding pulse G_x and the sampling period of the analogue to digital converter $\tau_{A/D}$, which is used to digitize the detected signal so that it could be fed to the computer for processing, where

$$\Delta_{k_x} = \frac{\gamma}{2\pi}G_x\tau_{A/D}. \tag{2.26}$$

For space limited images $f(x, y)$, overlap of the replications (known as aliasing) can be avoided by sampling at the Nyquist rate, which here corresponds to choosing the

parameters to make Γ just greater than the extent of the object.

One way of avoiding aliasing in the readout direction is by increasing the analogue to digital converter sampling time, i.e. reducing $\tau_{A/D}$, which increases the field of view according to Eqn 2.26. In the phase encoding direction aliasing is reduced by reducing Δ_{k_y}, which can be achieved by increasing the number of phase encodes, resulting in an increase in the scan time. Up to this point sampling has been considered to occur over an infinite extent, however in practice it can only occur over a finite extent. Therefore the sampled function over a finite extent W_{k_x} and W_{k_y} in the k_x and k_y directions respectively can be written as

$$\hat{F}_W(k_x, k_y) = F(k_x, k_y)S(k_x, k_y)\Pi(\frac{k_x}{W_{k_x}}, \frac{k_y}{W_{k_y}}), \tag{2.27}$$

where Π is the 2-D box function and

$$\begin{aligned} W_{k_x} &= 2(k_{xmax} + \frac{\Delta_{k_x}}{2}) \\ W_{k_y} &= 2(k_{ymax} + \frac{\Delta_{k_y}}{2}), \end{aligned} \tag{2.28}$$

where k_{xmax} and k_{ymax} are the maximum sampled frequencies in the k_x and k_y directions respectively. The inverse Fourier transform of $\hat{F}_W$ can be written as

$$\hat{f}_W(x, y) = f(x, y) \star\star s(x, y) \star\star W_{k_x} W_{k_y} sinc(W_{k_x} x) sinc(W_{k_y} y), \tag{2.29}$$

where $s(x, y)$ is the inverse Fourier transform of $S(k_x, k_y)$, $\star\star$ denotes 2-D convolution and $sinc(x) = \frac{sin(\pi x)}{\pi x}$. Therefore finite sampling in the frequency domain will result in blurring by a *sinc* function in the space domain. Spatial resolution δ_x and δ_y along the x and y directions respectively are defined as

$$\begin{aligned} \delta_x &= \frac{\Gamma_x}{N_{kx}} = \frac{1}{W_{kx}} \\ \delta_y &= \frac{\Gamma_y}{N_{ky}} = \frac{1}{W_{ky}}, \end{aligned} \tag{2.30}$$

where N_{kx} and N_{ky} are the number of data collected in the k_x and k_y directions respectively. It is obvious that the field of view and spatial resolution are related by the amount of data collected. Another important relationship is that between maximum frequencies sampled k_{xmax} and k_{ymax} in the k_x and k_y directions respectively and the pixel sizes Δ_x and Δ_y in the x and y directions respectively, which can be described by the following equation

$$\begin{aligned} \Delta_x &= \frac{1}{k_{xmax}} \\ \Delta_y &= \frac{1}{k_{ymax}}. \end{aligned} \tag{2.31}$$

To understand the effect of sampling on field of view, pixel size and resolution consider Fig. 2.12(a), which shows a representation of collected data from a MR scanner in the spatial frequency domain, and Fig. 2.12(b), which shows the corresponding data of the image in the space domain. If, for example, every other line of the raw data is removed as shown in Fig. 2.12(c), the field of view would be halved, according to Eqn 2.24, since both Δ_{k_x} and Δ_{k_y} are doubled. The spatial resolution will not change, as shown in Fig. 2.12(d), according to Eqn 2.30. However, if k_{xmax} and k_{ymax} are halved, as shown in Fig. 2.12(e), the pixel sizes are doubled and the spatial resolution is halved, whereas the field of view remains the same, as shown in Fig. 6.1(f).

2.5.4 Instrumental errors

The receiver of the MR system is very sensitive, since the MR signals are in the order of microvolts. External electromagnetic fields with frequencies (from radio and television stations, electric trains, CT scanners, type writers and floor cleaning appliances) within the frequency band of the receiver can be detected by the sensitive receiver and cause artifacts, which appear as streaks parallel to the phase encoding axis. The severity of the artifact depends on the strength of the source, how close the source is, strength of the RF shield and patient physique (tall patients act as antennas since more body is outside the magnet). Since the interfering signal generally has a fixed frequency, it corresponds to a fixed location on the frequency encoding axis while being independent of the phase encoding. The artifact can be minimized by carefully designing and testing the RF shield, monitoring the use of equipment (such as cleaning appliances) so that they are switched off during imaging and by modifying the magnetic field frequency until a clear band is found.

The direct current (DC) bright spot at the center is another artifact produced by instruments. Ideally a receiver produces no output signal when there is no input signal. In practical situations, the amplifier in the receiver will always produce some DC (zero spatial frequency) error, small constant output voltage, even in the absence of an input signal, which appears as a bright spot at the center of the reconstructed image. This artifact can be eliminated by the use of the 'phase alteration technique', where each exciting RF pulse is applied twice but with opposite polarity. By subtracting the two received signals, with opposite polarity, the DC effect is eliminated, while the required signal is doubled.

Ghost images and smearing, known as 'phase noise artifacts', in the phase encoding direction of the image are the most common artifacts noticed. One source of these artifacts is the pick up of unwanted low frequency noise by the gradient amplifiers that supply power to the gradient coils. Since MRI gradient amplifiers are very sensitive, they require noise precession by careful shielding and decoupling from external fields. Absence of power regulation to the RF transmitter amplifier may lead to fluctuation in

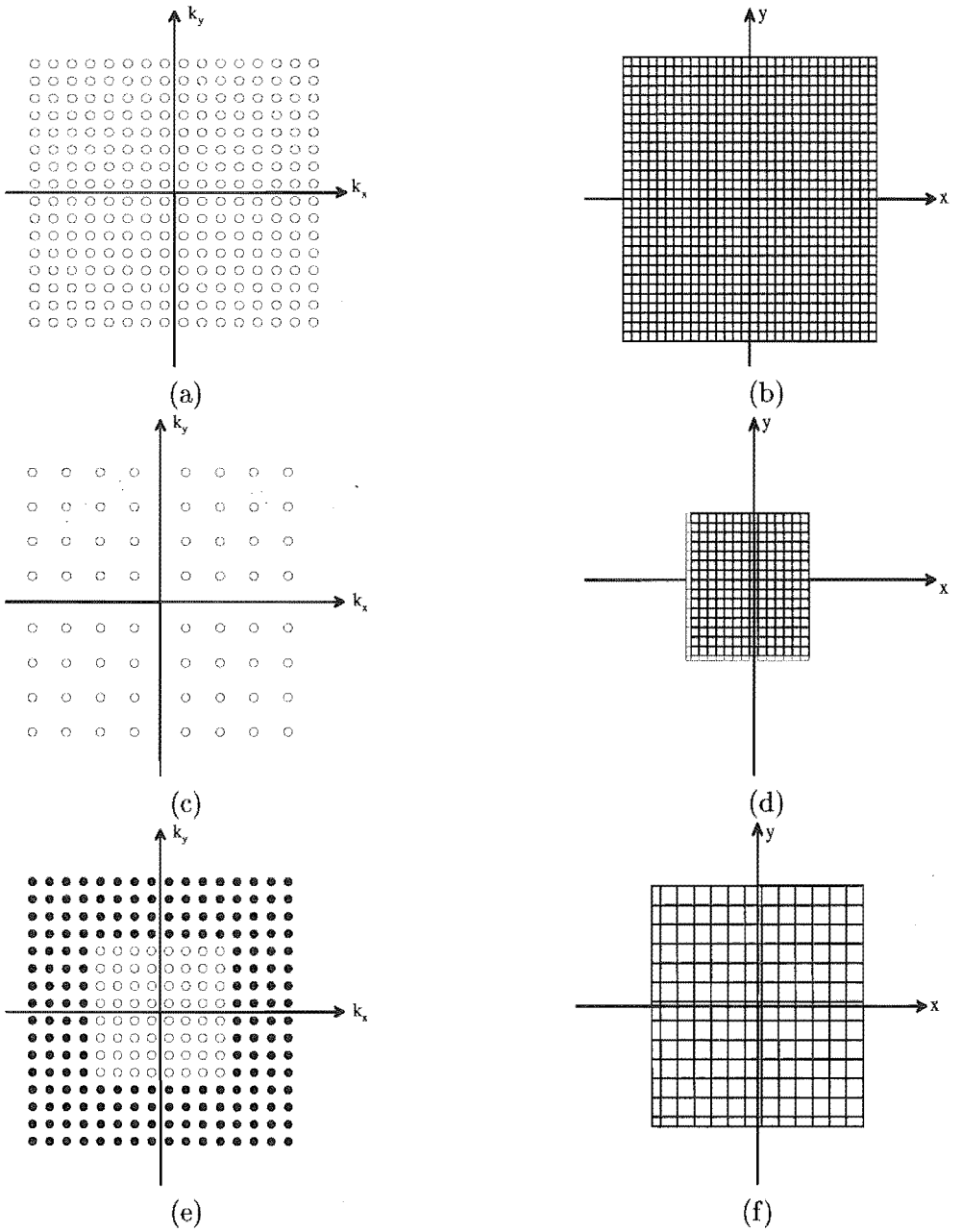


Figure 2.12 Relationships between sampling in the two domains, where a ○ represents a collected datum and • represents a missing datum which is set to zero: (a) raw data in k-space; (b) image space; (c) doubling Δ_{k_x} and Δ_{k_y} in k-space; (d) image space with field of view halved; (e) halving maximum sampled frequencies; (f) image space with pixel sizes doubled.

the input voltage to the amplifier, which is another source of the phase noise artifacts. This fluctuation results in unwanted alterations in the received MR signal, which appear as streaks parallel to the phase encoding axis [Atlas 1996].

2.5.5 Motion

The relatively long imaging time (in the order of minutes) of the conventional imaging methods means that artifacts due to motion need to be studied. Motion results in both ghosting and blurring of the MR images. There are several sources of motion such as

- random patient motion (especially when imaging children);
- periodic respiratory and cardiac motion; and
- fluid motion (e.g. blood).

Blurring due to motion is usually slight and in the direction of the motion. Ghost images, which are secondary images at locations different from the primary images, occur with a displacement in the phase encoding direction and are independent of the direction of the motion. The explanation for this phenomenon is that the data acquisition time along the frequency encoding direction lasts only a few milliseconds and displacement due to motion during this time is negligible. On the other hand, the interval over which phase encoding occurs is relatively long. There are many techniques to minimize or eliminate motion artifacts. Some of these techniques are described below [PHILIPS]

- Signal averaging: signals from several measurements of the same region of interest are summed. Since patient motion is random, averaging tends to reduce the artifact in the final image, but it results in an increase in imaging time.
- Triggering: this technique is used to synchronize data acquisition with motion. For cardiac motion, the data acquisition is synchronized by utilizing an electrocardiogram. Similarly, respiratory gating is used to generate a signal upon expiration.
- Breath holding: is not a very practical solution and is limited only to fast imaging techniques.

2.5.6 Noise

In the past few years a lot of work has been done with the aim of decreasing scan time and obtaining high resolution MR images. However the most fundamental restriction in achieving the above goals is the signal to noise ratio (SNR). Noise can be considered as a random component added to or subtracted from the voxel intensity. Since the

signal in a pixel (i.e. its brightness) is a function of the corresponding voxel intensity, therefore the SNR is an important criterion for the detectibility of region of interest [Felix *et al.* 1988]. It depends mainly on two parameters: 1) physical and instrumental parameters (e.g. applied magnetic field, size and shape of RF coils) and 2) imaging sequence parameters (e.g. data acquisition time, applied pulse sequence). The SNR in decibels is defined as [Nishimura 1994]

$$SNR = 20log_{10}\frac{\text{signal amplitude}}{\text{standard deviation of noise}}. \tag{2.32}$$

- Physical and instrumental parameters: The random noise in the MR system is due to ohmic losses in the receiving system. There are two sources for the losses: the ohmic losses in the RF coil and the losses from the body of the patient, where the latter is the dominant source of noise in a high field system.

  Assume an RF coil carrying a current I amperes, a field B_1 at point r is produced. Then if a transverse magnetization M_{xy} is placed at point r, a voltage V_o is induced in the coil given by

$$V_o(t) = \frac{-\delta}{\delta t}[M_{xy}]\beta_1 dV \tag{2.33}$$

  where $\beta_1 = \frac{B_1}{I}$ is the coil sensitivity, and dV is the volume of voxel considered. Since M_{xy} precess at the Larmor frequency, the amplitude of the signal is written as $V_o = \omega_0|M_{xy}||\beta_1|dV$. Both M_{xy} and ω_0 are $\propto B_0$, the signal amplitude is therefore proportional to B_0^2 [Vlaardingerbroek and Boer 1996].

  As mentioned above the main noise sources are the resistance of the RF coil (R_C) and the resistance of the patient as seen by the coil (R_P). Therefore, from the classical Nyquist equation, the noise power N_o can be written as

$$N_o = (4KT(R_C + R_P)\delta f)^{\frac{1}{2}} \tag{2.34}$$

  where K is Boltzman's constant, T is the absolute temperature and δf is the bandwidth of the receiver. So the SNR can be written as

$$SNR = \frac{\omega_0|M_{xy}||\beta_1|dV}{\sqrt{4KT(R_C + R_P)\delta f}} \tag{2.35}$$

  To find an expression for R_P, a simplified model shown in Fig. 2.13 is considered. A region with constant B_1 is assumed and a cylindrical conductive region is used as a model for the Eddy currents in the patient. The induced voltage $\mathcal{E}$ around

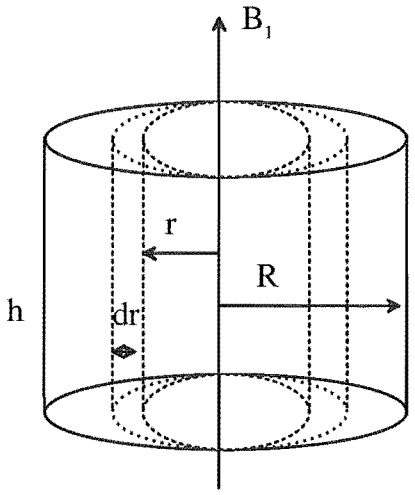


Figure 2.13 Model for calculation of losses due to patient's body.

the cylinder is [Hoult and Lauterbur 1979]

$$\mathcal{E} = -\frac{\delta\Phi}{\delta t} = -\int_{S_A} \frac{\delta B_1}{\delta t}.dS_A = j\omega_0 \pi r^2 B_1 \tag{2.36}$$

where Φ is the flux and S_A is the surface area. This induce voltage causes Eddy currents in a conducting medium. The conductance of a cylinder with conductivity σ, radius r and width δ_r is

$$\delta G = \frac{h\delta_r \sigma}{2\pi r} \tag{2.37}$$

Since the power dissipated $\delta P = \mathcal{E}^2 \delta G$, the total power P can be found by integrating δP over the whole volume. Assuming $B_1 = \beta_1$, i.e. $I = 1$, then $P = R_P$. Since $P \propto \omega_0^2 \Rightarrow R_P \propto B_0^2$.

The next noise source to consider is the RF coil resistance R_C, which is due to the skin effect in the copper wires used in the coil, which is given by

$$R_C = \frac{1}{\sigma \varrho}, \text{where} \tag{2.38}$$

$$\varrho = \frac{1}{\sqrt{\pi f \mu \sigma}} \tag{2.39}$$

where ϱ is the skin depth, μ is the permeability of the material and f is the frequency. Therefore $R_C \propto \frac{1}{\sqrt{\omega_0}} \propto \frac{1}{\sqrt{B_0}}$ [Hart *et al.* 1983]. The B_0 dependency of SNR could then be written as

$$SNR \propto \frac{B_0^2}{\sqrt{aB_0^{\frac{1}{2}} + bB_0^2}} \tag{2.40}$$

where a and b are the constants which account for the relative contributions to the noise from R_C and R_P respectively.

Another important parameter is the shape and size of the RF detection coil. While only the imaging slice is actively excited during an MR measurement, a much larger volume of the body is present in the detection volume of the coil. Therefore, surface coils that have a smaller sensitive volume provide a higher SNR compared to a body coil. Both the size of the body coil and its distance from the patient allows a greater contribution of noise from outside the selected slice as shown in Fig.. 2.14.

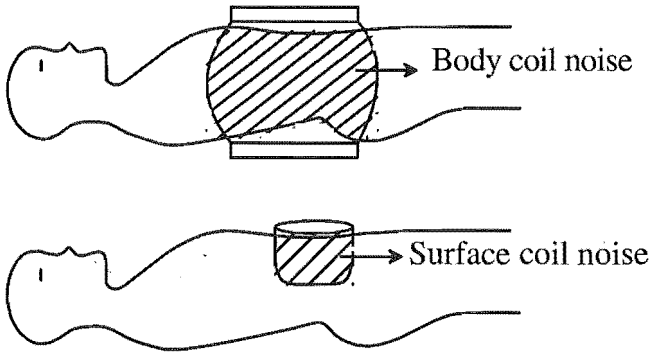


Figure 2.14 Effect of shape and size of coil on acquisition of background noise (a) body coil; (b) surface coil.

- Imaging sequence parameters: a reference SNR is first derived to study the effects of varying different parameters on the SNR. Assume an impulsive object that produces a signal with uniform amplitude A in the k-space. Therefore each of the N samples in the k-space will have an amplitude of value A and assuming Gaussian noise with zero mean and variance σ_n^2. Therefore, the SNR can be defined as [Nishimura 1994]

$$SNR = \frac{\sum_{j=1}^{N} A}{\sqrt{\sum_{j=1}^{N} \sigma_n^2}} = \frac{NA}{\sqrt{N\sigma_n^2}} = \frac{\sqrt{N}A}{\sigma_n} \tag{2.41}$$

One method of increasing SNR is signal averaging, which means that the signals are measured M times, and are then summed. The corresponding SNR is

$$SNR = \frac{\sum_{j=1}^{N} MA}{\sqrt{\sum_{j=1}^{N} M\sigma_n^2}} = \frac{NMA}{\sqrt{MN\sigma_n^2}} = \frac{\sqrt{MN}A}{\sigma_n}$$

Therefore, by increasing the number of times the image is measured from 1 to M and averaging, the SNR improves by $\sqrt{M}$. Doubling the sampling time also effects the SNR, since doubling the sampling interval corresponds to halving of the signal bandwidth, which results in halving of the noise variance (from Eqn.

2.34). Therefore,

$$SNR = \frac{NA}{\sqrt{N\frac{\sigma_n^2}{2}}} = \frac{\sqrt{2N}A}{\sigma_n}. \tag{2.42}$$

Another important parameter is the spatial resolution, which is determined by the voxel volume. Since the signal intensity is proportional to the voxel volume, halving the voxel volume(i.e. doubling the resolution) halves the signal intensity. To double the resolution, the readout gradient is doubled but both the sampling rate and readout time are unchanged. Since the receiver's bandwidth δf is inversely proportional to the sampling rate, keeping the sampling rate the same means the noise is unaffected and SNR is also halved.

2.5.7 Measurement errors

Another source of errors is the incorrect measurement of data or the measurement of undesired signals that are not related to the imaging process. Two different examples will be discussed below.

- Single digitization errors: to obtain a single image, there are at least 10^5 analogue to digital conversions performed. A single error in the measured data results in streaks that can be oriented at any angle and with spacing depending on the location of the erroneous data in the k-space. The erroneous data can usually be deleted, either by software or by the operator, and replaced by an interpolated value.

- Saturation of the analogue to digital (A/D) converter: in some images a halo appears in the background of the image and it may have an unusual contrast. The reason for this artifact is that the receiver attenuation is too low, resulting in a signal that exceeds the range of the A/D converter. This can be easily solved by readjusting the attenuation of the receiver for each measurement.

Chapter 3

SUPERRESOLUTION

3.1 INTRODUCTION TO SUPERRESOLUTION

The most important step in many imaging problems is the reconstruction of the object function f from the collected data F. Assuming 1-D functions, the object function $f(x)$ is reconstructed from $F(k)$ according to the following equation

$$f(x) = \int_{-\infty}^{\infty} F(k)e^{i2\pi kx}dk, \tag{3.1}$$

known as the inverse Fourier transform, which can be generalized to n-dimensions, with x representing a space coordinate and k the corresponding spatial frequency coordinate. To obtain $f(x)$ uniquely, if $f(x)$ is of finite support, $F(k)$ must be known over an infinite frequency range. However in many imaging problems, one cannot obtain all the data $F(k)$ to reconstruct $f(x)$. For example, in X-ray diffraction, phase that gives positional information is lost, while in astronomical imaging the finite aperture of the imaging system limits resolution. In medical and industrial imaging, complete sets of data are unobtainable due to practical reasons. For example, in fan beam CT imaging, limitation in the source motion restricts the collection of data at the optimal angular displacement, which is known as the limited view problem [Sezan and Stark 1984]. Simply setting the unknown data to zero causes significant artifacts in reconstructions (discussed in section 7.2). In MRI, where data is directly collected from a living human in the Fourier domain, the time needed to collect data is a limiting factor. The time must be minimized both to avoid motion artifacts and to maximize use of the expensive scanner. Typically the reconstructed image is blurred and suffers from Gibbs phenomenon. When only N discrete samples of information $F(k)$ are collected at $k = m\Delta k$ where $m = -\frac{N}{2} \ldots \frac{N}{2} - 1$ and $\Delta k = \frac{2\Omega}{N}$, where Ω is the maximum frequency available, $|k| \leq \Omega$. The reconstruction process for the discrete case is therefore

$$\hat{f}(n\Delta x) = \Delta k \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} G(m\Delta k)e^{i2\pi m\Delta k n\Delta x}, \tag{3.2}$$

$$\text{where} \qquad G(k) = \begin{cases} F(k) & |k| \leq \Omega \\ 0 & |k| > \Omega, \end{cases} \tag{3.3}$$

$\hat{f}(n\Delta x)$ is the reconstructed image, $n = -\frac{N}{2} \ldots \frac{N}{2} - 1$ and Δx is the sampling interval. To overcome the artifacts due to the truncated collected data, superresolution is required (explained below).

In this chapter, the idea of superresolution is briefly introduced followed by a proof of the achievability of superresolution by utilizing the concept of analytic continuation. A very simple and common method for superresolution known as the Gerchberg-Papoulis (G-P) algorithm is introduced in section 3.4 followed by a derivation of a mathematical formulation of the algorithm. The relationship between the G-P algorithm and the steepest descent method, used in solving linear systems, is investigated. Since the steepest descent method is very slow, the conjugate gradient method, presented in section 3.6, should be utilized since it is much faster and converges in a finite number of iterations. Finally, in section 3.7, results obtained by applying the above mentioned methods on two different sets of data are presented.

3.2 SUPERRESOLUTION

Superresolution, often referred to as bandwidth or spectral extrapolation, can be defined as the recovery of spatial frequency information of the object beyond the spatial frequency cutoff imposed by the transfer function of the imaging system [Hunt 1995]. The imaging process in 2-D space can be modelled by

$$G(k_x, k_y) \quad = \quad \mathbb{H}(k_x, k_y) F(k_x, k_y), \tag{3.4}$$

where G is the 2-D Fourier transform of the image (in MRI G is the collected data), $\mathbb{H}$ is the system transfer function of the image formation system, F is the 2-D Fourier transform of the imaged object and k_x, k_y are the spatial frequencies. The system transfer function determines the quality of the image. The limiting of the bandwidth is due to the property that

$$\mathbb{H}(k_x, k_y) = \begin{cases} \neq 0, & (k_x, k_y) \text{ inside region of support of } \mathbb{H} \\ = 0, & (k_x, k_y) \text{ outside region of support of } \mathbb{H}. \end{cases} \tag{3.5}$$

At a first glance one can conclude that by a simple rearrangement of Eqn. 3.4

$$F(k_x, k_y) \quad = \quad \frac{G(k_x, k_y)}{\mathbb{H}(k_x, k_y)}, \tag{3.6}$$

and therefore that superresolution is impossible since the only values of F that can apparently be recovered are those within the support region of $\mathbb{H}$, i.e. spatial frequencies outside the region of support of $\mathbb{H}$ are ambiguous. J. L. Harris proved however that

superresolution in theory is possible by using the theory of complex variables as follows [den Dekker and van den Bos 1997]. If $\mathbb{H}$ is known then F can be recovered within the passband of the imaging system. If the imaged object is finite in size, then its spectrum is analytic. The complete and exact spectrum F can then be found by analytic continuation of the known part of F, explained below.

3.3 ANALYTIC CONTINUATION

Prolate spheroidal wave functions can be used to analytically continue an image spectrum beyond its diffraction limit. In 1-D space, there is an infinite set of real functions $\psi_i(x)$ and real numbers κ_i with the following properties for given real numbers $X > 0$ and $\Omega > 0$ [Slepian and Pollack 1961]:

- The spectrum of $\psi_i(x)$ is zero for $|k| > \Omega$.
- $\psi_i(x)$ are orthonormal on the real line

$$\int_{-\infty}^{\infty} \psi_i(x)\psi_j(x)dt \quad = \quad \delta(i-j). \tag{3.7}$$

- $\psi_i(x)$ are orthogonal in the interval $\frac{-X}{2} \leq x \leq \frac{X}{2}$

$$\int_{-\frac{X}{2}}^{\frac{X}{2}} \psi_i(x)\psi_j(x)dt \quad = \quad \kappa_i\delta(i-j). \tag{3.8}$$

- For all values of x, real or complex

$$\kappa_i\psi_i(x) \quad = \quad \int_{-\frac{X}{2}}^{\frac{X}{2}} \frac{\sin\Omega(x-s)}{\pi(x-s)}\psi_i(s)ds, \tag{3.9}$$

  for $i, j = 0, 1, 2\ldots$

where $\psi_i(x)$ are functions which depend on X and Ω. In 2-D space, suppose $f(x, y)$ is confined in space such that

$$f(x,y) \quad = \quad 0 \qquad \text{for } |x| > \Gamma \quad \text{and} \quad |y| > \Gamma, \tag{3.10}$$

and assume that the spectrum $F(k_x, k_y)$ of $f(x, y)$ is known only over a certain spatial frequency range defined by

$$\begin{aligned} |k_x| &\leq k_{xc} \\ |k_y| &\leq k_{yc}. \end{aligned} \tag{3.11}$$

$F(k_x, k_y)$ could be written in terms of the prolate spheroidal functions $\psi_i(k)$ such that [Pratt 1978]

$$F(k_x, k_y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_{ij} \psi_i(k_x) \psi_j(k_y). \tag{3.12}$$

The coefficients a_{ij} can be found by multiplying both sides of Eqn. 3.12 by $\psi_m(k_x)\psi_n(k_y)$, integrating over the passband and using Eqn. 3.8 such that

$$a_{mn} = \frac{1}{\kappa_m \kappa_n} \int_{-k_{xc}}^{k_{xc}} \int_{-k_{yc}}^{k_{yc}} \psi_m(k_x)\psi_n(k_y) F(k_x, k_y) dk_x dk_y, \tag{3.13}$$

for $m, n = 0, 1, 2 \ldots$ Substituting for the a_{ij} in Eqn. 3.12 using Eqn. 3.13, gives an expansion of $F(k_x, k_y)$ over all values of k_x and k_y, implying that $F(k_x, k_y)$ is known everywhere. While analytic continuation is theoretically possible, however, practically it is very difficult to perform. Reasons include the truncation errors that arise due to the finite implementation of Eqn. 3.12, and the presence of noise in the known part of the spectrum $F(k_x, k_y)$, resulting in an inaccurate evaluation of the coefficients according to Eqn. 3.13.

3.4 THE GERCHBERG-PAPOULIS (G-P) ALGORITHM

One of the earliest and simplest algorithms available for superresolution is the Gerchberg-Papoulis algorithm. This algorithm is based on two principles:

- Recovery of object information beyond the limited bandwidth is achieved by utilizing information within the limited bandwidth.
- Knowledge of the object size is utilized in recovery of information beyond the limited bandwidth.

The G-P algorithm is an iterative algorithm which is based on the error energy reduction principle. Fig. 3.1 is a block diagram illustrating the G-P algorithm. To superresolve an image using the G-P algorithm, the image's spectrum G is first inverse Fourier transformed to obtain $\tilde{f}_{GP}^0$. The space domain constraints are then applied on the image $\tilde{f}_{GP}^0$, by setting the object to zero beyond the support region. The frequency domain constraints are then applied on the modified image's spectrum, by correcting the spectrum within the limited bandwidth [Gerchberg 1974, Papoulis 1975]. The above process is iterated several times until a desired criterion is achieved. The following sections will study the G-P algorithm and prove how it achieves superresolution.

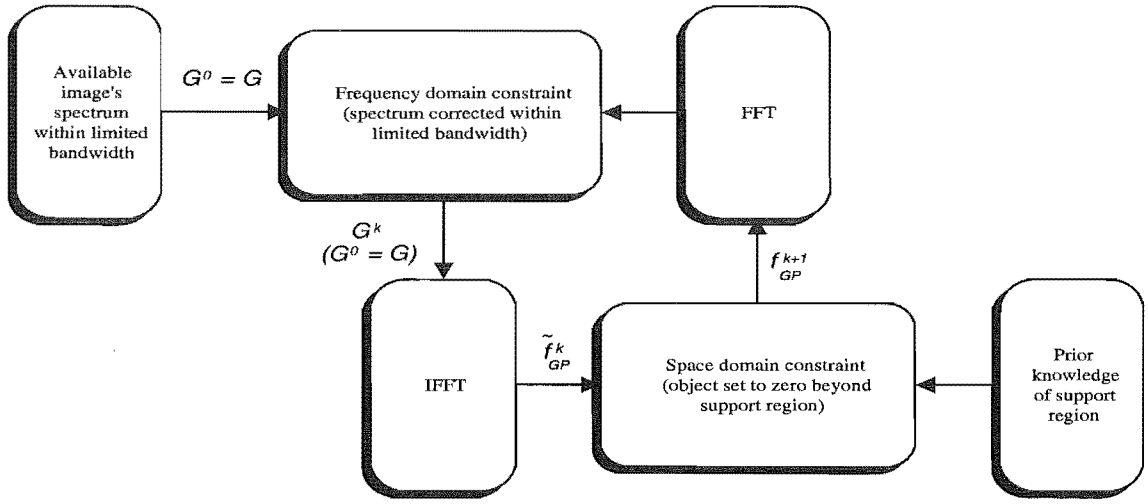


Figure 3.1 A block diagram illustrating G-P algorithm.

3.4.1 Mathematical formulation of the G-P algorithm

The G-P algorithm can be mathematically formulated as an iterative equation. The space domain constraint $\gamma(x)$ with a support limit X can be written as

$$\gamma(x) = \begin{cases} 1 & |x| \leq X \\ 0 & |x| > X, \end{cases} \tag{3.14}$$

and the frequency domain constraint $\xi(k) = 1 - \mu(k)$ with a bandlimit Ω can be written as

$$\xi(k) = \begin{cases} 1 & |k| \leq \Omega \\ 0 & |k| > \Omega. \end{cases} \tag{3.15}$$

Using Fig. 3.1 and assuming the initial estimate of the spectrum of the true object $f(x)$ is G (i.e. $G^0 = G$), we have

$$\begin{aligned}
\tilde{f}^0_{GP} &= \mathcal{F}^{-1}G^0 \\
f^1_{GP} &= \gamma\tilde{f}^0_{GP} = \gamma\mathcal{F}^{-1}G^0 \\
G^1 &= \mu\mathcal{F}f^1_{GP} + G^0 \\
\tilde{f}^1_{GP} &= \mathcal{F}^{-1}G^1 \\
f^2_{GP} &= \gamma\tilde{f}^1_{GP} = \gamma\mathcal{F}^{-1}(\mu\mathcal{F}f^1_{GP} + G^0) \\
&\vdots \\
G^k &= \mu\mathcal{F}f^k_{GP} + G^0 \qquad (3.16) \\
\tilde{f}^k_{GP} &= \mathcal{F}^{-1}G^k \qquad (3.17) \\
f^{k+1}_{GP} &= \gamma\tilde{f}^k_{GP} \qquad (3.18) \\
&\vdots
\end{aligned}$$

where f_{GP}^k is the recovered image at the k^{th} iteration (after applying the spatial domain constraints), and $\mathcal{F}$ and $\mathcal{F}^{-1}$ are forward and inverse Fourier transforms respectively. Therefore, G-P algorithm at the $k+1$ iteration can be written as [Connolly *et al.* 1995]

$$f_{GP}^{k+1} = \gamma \mathcal{F}^{-1}(\mu \mathcal{F} f_{GP}^k + G). \tag{3.19}$$

Another approach to proving that the G-P algorithm solves Eqn. 3.4 is by relating it to the steepest descent method, introduced in the next section, which solves linear systems. It will be proven that the G-P algorithm is equivalent to the steepest descent method for a specific steplength value.

3.4.2 G-P algorithm performance in a noise free case

To understand the way the G-P algorithm superresolves an image, consider, in the noise free case, the image spectrum $G(k)$ to be the sum of the true object spectrum $T_r(k)$ and the error spectrum $E_r(k)$ as illustrated in Fig. 3.2.

To superresolve an image, $E_r(k)$ should be reduced so that $G(k)$ approaches $T_r(k)$. Since the G-P algorithm is a linear process, its effect on $T_r(k)$ and $E_r(k)$ can be studied independently. Since $E_r(k)$ has a finite region of zeros, it cannot be an analytic function (analytic functions may only have isolated zeros), so its Fourier transform extends to infinity. Applying the space domain constraint on the Fourier transformed

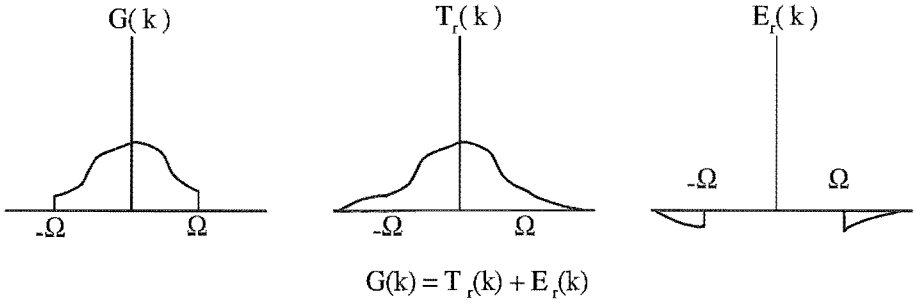


Figure 3.2 Relation between image spectrum $G(k)$, true spectrum $T_r(k)$ and error spectrum $E_r(k)$.

energy spectrum must therefore reduce the energy of the transformed error function. A reduction of energy in the space domain is equivalent to a reduction of energy in the frequency domain by Parseval's theorem. Now the modified error function is finite, so its spectrum is analytic and therefore will contain some energy within the original bandlimit. Applying the frequency domain constraints on the spectrum of the modified error function (i.e. reducing the error inside the bandlimit to zero) results in a further reduction in the error energy. Therefore in each iteration the energy of the error spectrum is reduced twice in succession. The G-P algorithm is therefore correctly referred to as the error reduction algorithm.

3.4.3 G-P algorithm performance in a noisy case

It is important to study the effect of the G-P algorithm on noise, since in practice collected data is always contaminated by noise. Referring to Eqn. 3.16, we have

$$G^{k+1} = \mu \mathcal{F} f_{GP}^{k+1} + G^0. \tag{3.20}$$

Substituting Eqns. 3.17 and 3.18 into Eqn. 3.20, we have

$$G^{k+1} = \mu \mathcal{F}(\gamma \mathcal{F}^{-1} G^k) + G^0, \tag{3.21}$$

where $G^0 = T_r + N_o$, where T_r is the true object spectrum and N_o is the noise. Since the G-P algorithm is a linear process, therefore the effect of the G-P algorithm on noise is [Connolly *et al.* 1995]

$$N_o^{k+1} = \mu \mathcal{F}(\gamma \mathcal{F}^{-1} N_o^k) + N_o. \tag{3.22}$$

Considering the best case scenario when most of the object spectrum is measured (i.e. $\mu \approx 0$), Eqn. 3.22 becomes

$$N_o^{k+1} = N_o, \tag{3.23}$$

indicating that in every iteration of the G-P algorithm, noise is neither reduced nor increased. Therefore the application of the G-P on a noisy image when $\mu \approx 0$ results in a superresolved image with the same amount of noise as in the collected data. For the worst case scenario, when very little data is collected (i.e. $\mu \approx 1$) and when $\gamma = 1$ we have

$$N_o^{k+1} = N_o^k + N_o = \sum_{i=0}^{k+1} N_o = (k+1)N_o, \tag{3.24}$$

implying that noise is amplified at each iteration. In practice μ and γ are usually in between the above extreme cases indicating that noise is amplified at each iteration i.e. $N_o^{k+1} = N_o + \sum_{i=1}^{k+1} CN_o$ where the constant $C < 1$. The strong effect of noise is due to the ill-posed nature of the problem. Regularization methods, discussed in chapter 5, aim to stabilize the problem and make it less sensitive to noise.

3.5 STEEPEST DESCENT METHOD

The steepest descent method is one of the oldest methods available for solving the following linear system [Shewchuck 1994]

$$A\zeta = b, \tag{3.25}$$

where ζ is an unknown $n \times 1$ vector, b is a known $n \times 1$ vector and A is a known, $n \times n$ square symmetric, positive-definite matrix, such that $\zeta^T A\zeta$ is positive for nonzero values of ζ, where T indicates the transpose of a matrix. In order to solve Eqn. 3.25, consider first a quadratic function $\rho(\zeta)$ which is simply defined as

$$\rho(\zeta) = \frac{1}{2}\zeta^T A\zeta - b^T\zeta + c, \tag{3.26}$$

where $\rho(\zeta)$ is a scalar quadratic function of the vector ζ, and c is a constant. Since A is a positive-definite function, $\rho(\zeta)$ has a paraboloid bowl surface and at the bottom of the paraboloid bowl, the gradient is zero. The gradient of $\rho(\zeta)$ is defined as

$$\rho'(\zeta) = \begin{bmatrix} \frac{\delta}{\delta\zeta_1}\rho(\zeta) \\ \frac{\delta}{\delta\zeta_2}\rho(\zeta) \\ \vdots \\ \frac{\delta}{\delta\zeta_n}\rho(\zeta) \end{bmatrix}, \tag{3.27}$$

where the gradient of $\rho(\zeta)$ at point ζ, indicates the direction of the greatest increase in $\rho(\zeta)$. Therefore, to minimize $\rho(\zeta)$, the gradient is set to zero. Calculating the gradient of Eqn. 3.26, we obtain

$$\rho'(\zeta) = \frac{1}{2}A^T\zeta + \frac{1}{2}A\zeta - b, \tag{3.28}$$

which reduces to $\rho'(\zeta) = A\zeta - b$, if A is symmetric. Thus $A\zeta = b$ can be solved by finding a ζ that minimizes $\rho(\zeta)$.

To illustrate this, consider an example where

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ -8 \end{bmatrix}, \quad c = 0. \tag{3.29}$$

Fig. 3.3(a),(b) illustrate the system $A\zeta = b$ and the corresponding quadratic form $\rho(\zeta)$ respectively, where the intersection of the 2 lines correspond to the minimum point of the paraboloid.

The steepest descent method starts with an arbitrary point ζ^0 and then steps to another point ζ^1 in such a way that the step direction is toward the most rapid decrease in $\rho(\zeta)$. This method iterates until a value of ζ^k sufficiently close to the minimum of $\rho(\zeta)$ is found. In other words, the iterative method can be written as

$$\zeta^{k+1} = \zeta^k + \alpha^k r^k, \tag{3.30}$$

$$\text{where } r^k = -\rho'(\zeta^k)$$

$$= b - A\zeta^k, \tag{3.31}$$

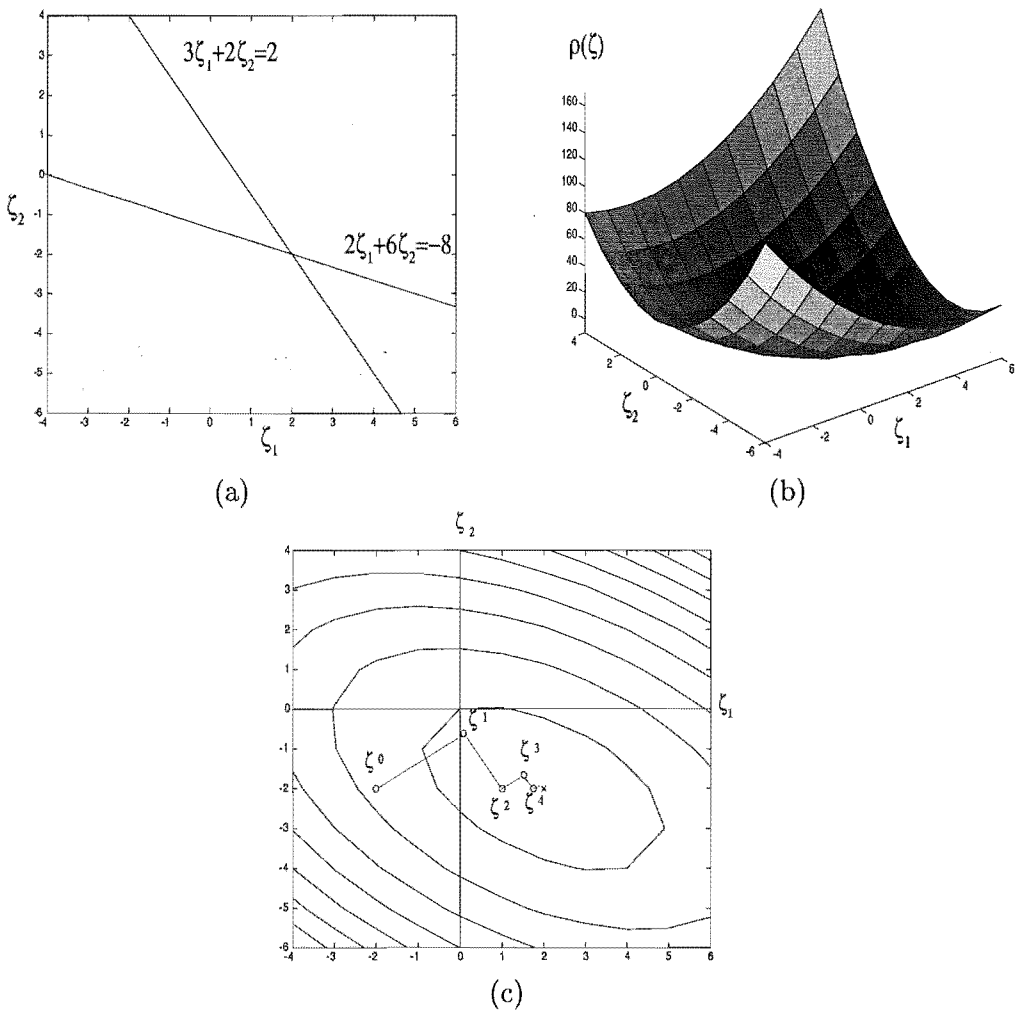


Figure 3.3 Steepest descent example: (a) $A\zeta = b$; (b) $\rho(\zeta)$; (c) path of the steepest descent method along the contours of $\rho(\zeta)$), with $\times$ indicating the solution.

is the direction of the steepest descent, and where α^k is the steplength, which minimizes ρ when

$$\begin{aligned}
\frac{d}{d\alpha^k}\rho(\zeta^{k+1}) &= 0 \Rightarrow \\
\rho'(\zeta^{k+1})^T \frac{d\zeta^{k+1}}{d\alpha^k} &= \rho'(\zeta^{k+1})^T r^k = 0 \quad \text{by the chain rule} \\
\rho'(\zeta^{k+1})^T r^k &= (b - A\zeta^{k+1})^T r^k \\
&= (b - A(\zeta^k + \alpha^k r^k))^T r^k = 0 \Rightarrow \\
(b - A\zeta^k)^T r^k &= \alpha^k (Ar^k)^T r^k \Rightarrow \\
r^{kT} r^k &= \alpha^k (r^k)^T (Ar^k) = 0 \Rightarrow \\
\alpha^k &= \frac{(r^k)^T r^k}{(r^k)^T Ar^k}.
\end{aligned} \tag{3.32}$$

For the example considered above, Fig. 3.3(c) shows the the path traversed by applying the steepest descent method with the starting point $\zeta^0 = [-2, -2]^T$ until it converges to the solution $[2, -2]^T$.

To relate the steepest descent method to the G-P algorithm, consider a spatially bounded object $f(x)$ such that $f(x) = 0$ for $|x| > X$, and $G(k)$ is the bandlimited spectrum obtained through the imaging process, such that $G(k) = 0$ for $|k| > \Omega$. The superresolution problem can be written as [Connolly *et al.* 1995, Connolly and Lane 1997]

$$\begin{aligned}
G &= Hf \\
&= \xi \mathcal{F} \gamma f,
\end{aligned} \tag{3.33}$$

where $\xi, \mathcal{F}$ and γ are as defined previously. The aim of superresolution is to find f such that $\|\xi \mathcal{F} \gamma f - G\|^2$ is minimized, where

$$\begin{aligned}
\|f\|^2 &= < f, f > \quad \text{and is known as the Euclidean norm and} \\
< f, h > &= \int_a^b f(x)h(x)dx \quad \text{is known as the inner product.}
\end{aligned} \tag{3.34}$$

Comparing the superresolution problem to the steepest descent approach we have

$$f^{k+1} = f^k + \alpha^k r^k, \tag{3.35}$$

where

$$\begin{aligned} r^k &= -\frac{\delta}{\delta f^k}\|\xi\mathcal{F}\gamma f^k - G\|^2 \\ &= -\frac{\delta}{\delta f^k}[< f^k, \gamma\mathcal{F}^{-1}\xi\mathcal{F}\gamma f^k > -2 < f^k, \gamma\mathcal{F}^{-1}\xi G > + < G, G >] \\ &= 2\gamma\mathcal{F}^{-1}\xi G - 2\gamma\mathcal{F}^{-1}\xi\mathcal{F}\gamma f^k \\ &= 2\gamma\mathcal{F}^{-1}G - 2\gamma\mathcal{F}^{-1}\xi\mathcal{F}f^k. \end{aligned} \tag{3.36}$$

The above equation was derived using the following properties

$$\begin{aligned} < Hf^k, Hf^k > &= < f^k, H^\dagger Hf^k > \\ H &= \xi\mathcal{F}\gamma \\ \Rightarrow H^\dagger &= \gamma\mathcal{F}^{-1}\xi \\ \xi G &= G \quad \text{since } G \text{ is bandlimited} \\ \gamma f^k &= f^k \quad \text{since } f^k \text{ has a finite support.} \end{aligned} \tag{3.37}$$

where $H^\dagger$ implies the adjoint of H. Substituting Eqn. 3.36 in Eqn. 3.35, and using $\xi = 1 - \mu$ and setting $\alpha^k = \frac{1}{2}$ we have

$$f^{k+1} = \gamma\mathcal{F}^{-1}(\mu\mathcal{F}f^k + G), \tag{3.38}$$

which is the same as Eqn. 3.19. Therefore, the G-P algorithm is equivalent to the steepest descent method, with $\alpha^k = \frac{1}{2}$.

3.6 CONJUGATE GRADIENT (C-G) METHOD

The conjugate gradient method is used to solve the problem of minimizing quadratic functions [Luenberger 1984]. The method proceeds by generating vector sequences of iterates (i.e. successive approximations to the solution), residuals corresponding to the iterates, and search directions used in updating the iterates and residuals. The iterative approximations to the solution can be written as

$$\zeta^{k+1} = \zeta^k + \alpha^k d^k, \tag{3.39}$$

where $d^0, d^1 \ldots d^{n-1}$ is a set of orthogonal search directions. The search directions d^k are chosen such that they are *A-orthogonal* to each other and to e^{k+1} (i.e. $d^{iT}Ad^j = 0, i \neq j$ and $d^{kT}Ae^{k+1} = 0$), where $e^{k+1} = \zeta^{k+1} - \zeta$, with ζ being the solution. Setting the

directional derivative to zero and using the chain rule we have

$$\begin{aligned} \frac{d}{d\alpha^k}\rho(\zeta^{k+1}) &= 0 \Rightarrow \\ -r^{k+1^T} d^k &= 0 \Rightarrow \\ d^{k^T} A e^{k+1} &= 0, \end{aligned} \tag{3.40}$$

where the latter part of the above equation was derived using $r^k = b - A\zeta^k = b - A(e^k + \zeta) = -Ae^k$. The derivation of α^k can be done in the same way as Eqn. 3.32 to obtain:

$$\alpha^k = \frac{d^{k^T} r^k}{d^{k^T} A d^k}. \tag{3.41}$$

The set of *A-orthogonal* search directions can be found using the conjugate Gram-Schmidt method where

$$d^i = u^i + \sum_{k=0}^{i-1} \beta_{ik} d^k \qquad \text{for} \quad i > 0, \tag{3.42}$$

where u is a set of n linearly independent vectors. For the conjugate gradient method we have $u^k = r^k$, and $d^0 = u^0$. The values of β_{ik} can be computed using

$$\begin{aligned} d^{i^T} A d^j &= u^{i^T} A d^j + \sum_{k=0}^{i-1} \beta_{ik} d^{k^T} A d^j \\ &= 0 \Rightarrow \\ \beta_{ij} &= -\frac{u^{i^T} A d^j}{d^{j^T} A d^j}. \end{aligned} \tag{3.43}$$

The disadvantage of this method is that all the old search vectors should be saved in memory for computation of β_{ij}, however this can be overcome by simplifying the above equation. A few derivations should be made first

$$\begin{aligned} r^{k+1} &= -Ae^{k+1} \\ &= -A(e^k + \alpha^k d^k) \\ &= r^k - \alpha^k A d^k. \end{aligned} \tag{3.44}$$

The following identities should also be stated

$$\begin{aligned} d^{i^T} r^j &= 0 \qquad \text{for} \quad i < j \\ d^{i^T} r^i &= u^i r^i \\ r^{i^T} r^j &= 0 \qquad \text{for} \quad i \neq j. \end{aligned} \tag{3.45}$$

Using the above derivations, β_{ij} can be simplified as follows

$$\begin{aligned} r^{iT}r^{j+1} &= r^{iT}r^{j} - \alpha^{j} r^{iT} A d^{j} \\ \alpha^{j} r^{iT} A d^{j} &= r^{iT}r^{j} - r^{iT}r^{j+1}, \end{aligned} \tag{3.46}$$

$$r^{iT} A d^{j} = \begin{cases} \frac{r^{iT}r^{i}}{\alpha^{i}}, & i = j \\ -\frac{r^{iT}r^{i}}{\alpha^{i-1}}, & i = j+1 \\ 0, & otherwise, \end{cases} \tag{3.47}$$

$$\begin{aligned} \Rightarrow \beta_{i,i-1} &= \beta_i \\ &= \frac{r^{iT}r^{i}}{r^{i-1T}r^{i-1}}. \end{aligned} \tag{3.48}$$

The conjugate gradient method can then be summarised as

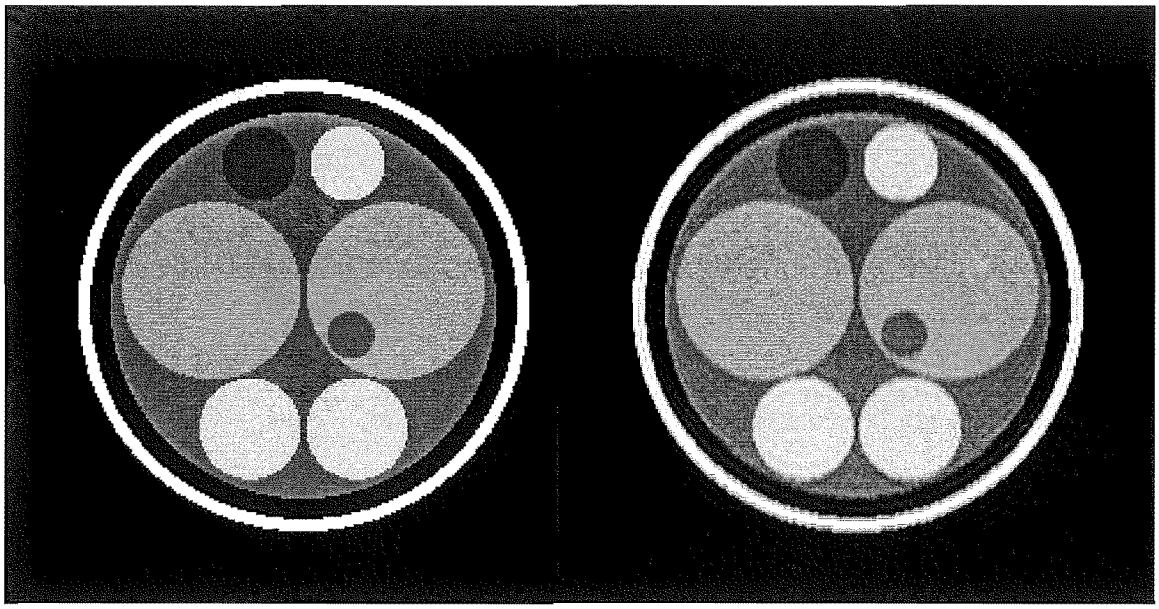
$$\begin{aligned} d^{0} &= r^{0} \\ &= b - A\zeta^{0} \\ \alpha^{i} &= \frac{r^{iT}r^{i}}{d^{iT}Ad^{i}} \\ \zeta^{i+1} &= \zeta^{i} + \alpha^{i} d^{i} \\ r^{i+1} &= r^{i} - \alpha^{i} d^{i} \\ \beta_{i+1} &= \frac{r^{i+1T}r^{i+1}}{r^{iT}r^{i}} \\ d^{i+1} &= r^{i+1} + \beta_{i+1} d^{i}. \end{aligned} \tag{3.49}$$

The first iteration of both the steepest descent and the conjugate gradient methods is the same, but the next step advances in a direction that is a linear combination of the current residual and the previous direction vector. Even though the conjugate gradient method is slightly more complicated than the steepest descent method, it is faster and it converges in a finite number of steps.

3.7 RESULTS

A simple 'head' phantom was simulated by superimposing several circular disks of uniform amplitude. The resulting object had a circular support of 190 pixels diameter within a 256 × 256 pixel region and 5 distinct nonzero amplitude levels (in range 0 to 1) as shown in Fig. 3.4(a), and is referred to as the high resolution head phantom. The DFT of the original phantom was bandlimited by truncating the spectrum to the lowest 101 spatial frequencies vertically and the lowest 201 frequencies horizontally. Each of the spatial components within the bandlimit was corrupted by complex pseudorandom

Gaussian noise so that the overall SNR within the bandlimit was $30dB$. The inverse DFT of the noisy bandlimited spectrum is shown in Fig. 3.4(b), and is referred to as bandlimited head image. Note the ringing caused by the truncation, especially adjacent to the horizontal edges.



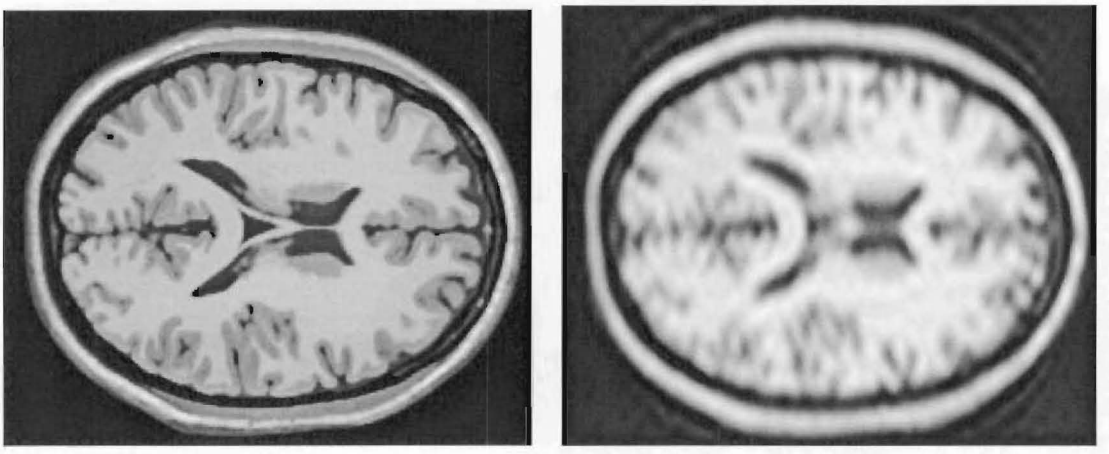
(a) (b)

Figure 3.4 (a) High resolution head phantom; (b) bandlimited noisy head image.

A second image was obtained from an MRI simulator available on the World Wide Web (WWW) [Kwan *et al.* 1996]. The simulator simulates the application of a pulse sequence with configurable parameters to a 3-D brain model, generated from a volumetric MRI data set, to generate the MRI volume [Cososco *et al.* 1997, Collins *et al.* 1998]. One 2-D slice, with dimension 180×216, was selected as shown in Fig. 3.5(a) and is referred to as the high resolution brain image. The image was blurred by bandlimiting the spectrum to the lowest 31 spatial frequencies vertically and the lowest 71 spatial frequencies horizontally. Each of the spatial components within the bandlimit was corrupted by complex pseudorandom Gaussian noise so that the overall SNR within the bandlimit was $30dB$. The inverse DFT of the noisy bandlimited spectrum is shown in Fig. 3.5(b) and is referred to as the bandlimited brain image. Both the G-P algorithm and the C-G method are applied on both data sets. The criteria used to study the performance of the different methods is the improvement in signal to noise ratio (ISNR). The ISNR is defined as [Banham and Katsaggelos 1997]

$$ISNR = 10\log_{10}\frac{\sum(f-g)^2}{\sum(f-\hat{f})^2}, \tag{3.50}$$

where $\hat{f}$ is the recovered image, f is the original object image (high resolution image),



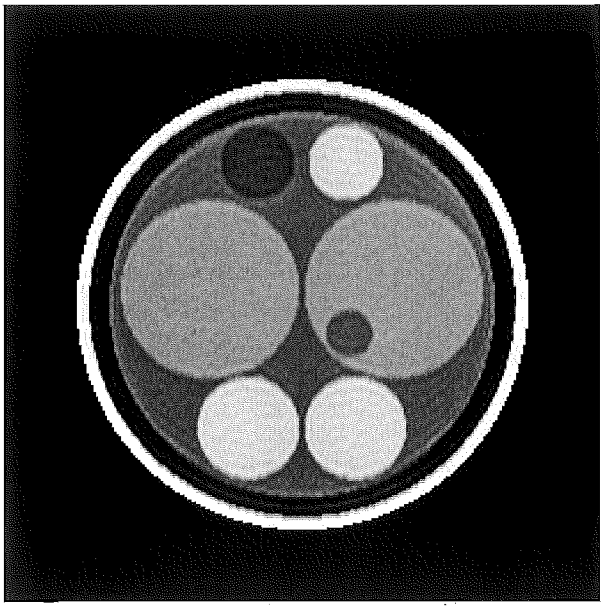
(a) (b)

Figure 3.5 (a) High resolution brain image; (b) bandlimited noisy brain image.

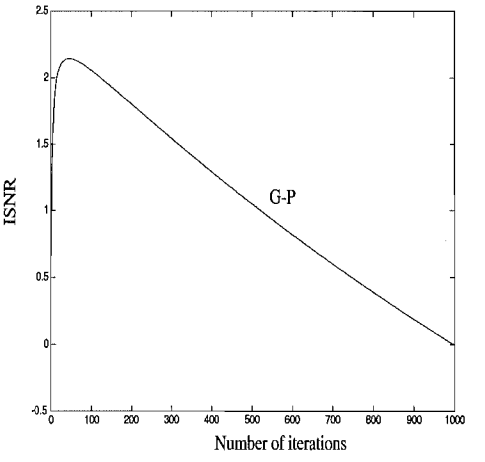
and g is the bandlimited noisy image (low resolution image), and the summations are over all pixels within the support. It is important to note that the ISNR by itself is not enough to judge the performance of the methods. A visual comparison between the bandlimited images and their corresponding recovered images is also very important.

The G-P algorithm was first applied to the bandlimited head image; Fig. 3.6(a) shows the recovered image and Fig. 3.6(b) shows the corresponding ISNR curve. The recovered image was obtained at iteration 82 where the ISNR had the maximum value. There is a very slight improvement in the recovered image compared to the bandlimited image. One of the problems of the G-P algorithm is that its performance (ISNR) increases to a maximum and then starts to decrease as the number of iterations increase, due to the nonconvergence of the algorithm in the presence of noise. Chapter 5 discusses various methods which can be incorporated within the G-P algorithm to level off the ISNR curve so that it is no longer sensitive to the number of iterations performed. The C-G method was also applied on the bandlimited head image, Fig. 3.7(a) shows the recovered image obtained at the iteration for which the ISNR was at its maximum value, and Fig. 3.7(b) shows the corresponding ISNR curve. The C-G method is faster than the G-P algorithm where the ISNR curve reaches the maximum at iteration number 15 whereas in the G-P algorithm it reaches it at iteration number 82. Note that consequent divergence of the C-G method is also however more rapid than that for the G-P algorithm. In the rest of the thesis, we restrict ourselves to iterative methods based on the G-P or steepest descent methods.

The G-P algorithm was applied on the bandlimited brain image, Fig. 3.8(a) shows the recovered image (obtained at iteration 6) and Fig. 3.8(b) shows the corresponding ISNR curve obtained using the G-P algorithm. Again there is a minimal visual improvement in the recovered image compared to the bandlimited brain image, which indicates a limitation in the amount of superresolution obtainable when only the sup-

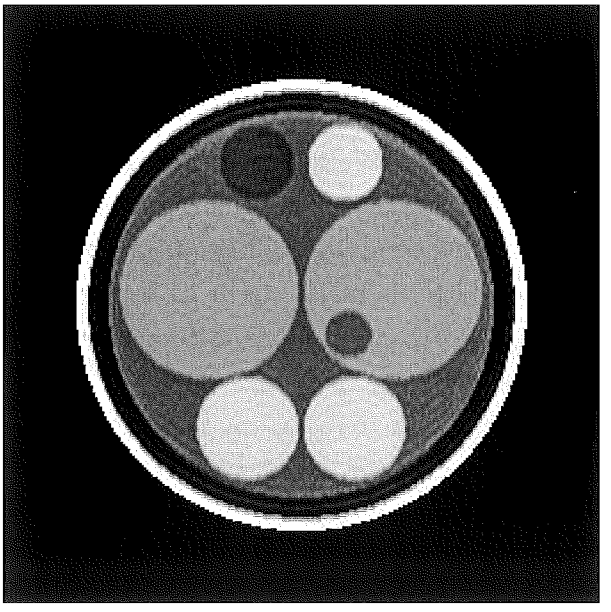


(a)

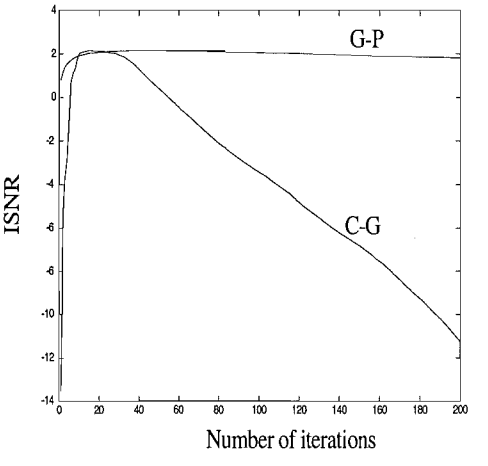


(b)

Figure 3.6 G-P algorithm: (a) recovered head image; (b) ISNR curve.

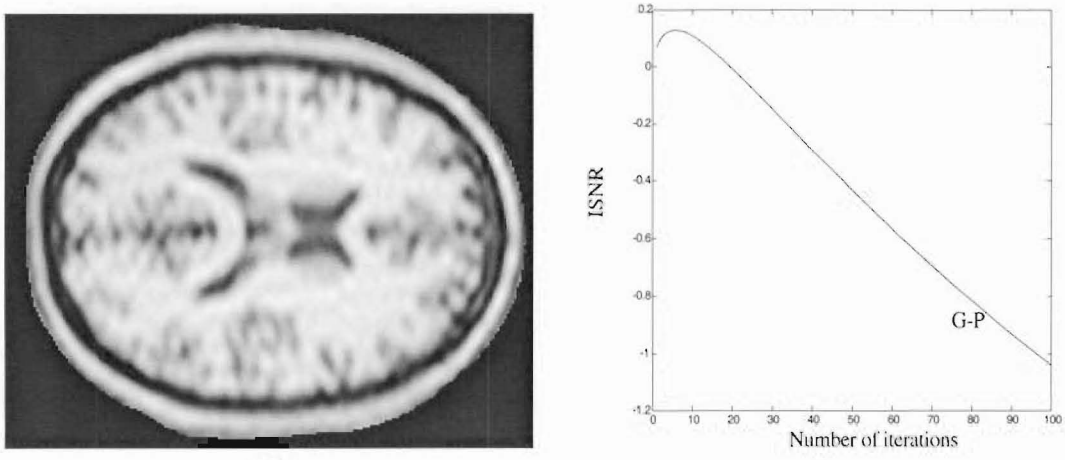


(a)



(b)

Figure 3.7 C-G algorithm: (a) recovered head image; (b) ISNR curve.



(a) (b)

Figure 3.8 G-P algorithm: (a) recovered brain image; (b) ISNR curve.

port constraint is used. Additional constraints will be considered in the next chapter to increase the amount of superresolution obtained.

Chapter 4

PROJECTION ONTO CONVEX SETS (POCS)

4.1 INTRODUCTION TO POCS

In the past few decades, much work has been done on reconstructing images from incomplete data. The reconstruction process is used in various applications such as reconstruction of tomographic images from incomplete view data [Sezan and Stark 1984], spectrum extrapolation, restoring images when only magnitude information is available, etc. In many applications such as CT or MRI, it may be impossible to measure a complete set of data. For example in medical CT applications, obtaining a complete set of data means increasing dosage and scanning time which is unrealistic [Sezan and Stark 1983]. Similarly in MRI, obtaining a complete set of data means keeping the patient still for a long period of time. Hence the G-P and other similar algorithms have a significant role in the area of superresolution. A more general and recent algorithm known as projection onto convex sets (POCS), developed by Youla and Webb [Youla and Webb 1982] and popularized by Stark [Stark 1987], has received wide interest. POCS is a simple, fast iterative technique which utilizes prior information about the imaged object in the restoration process. It uses each item of prior information to form a constraint and seeks an image that lies in the intersection of all the constraints.

This chapter starts with a simple overview of set theoretic estimation, convex sets and their properties, followed by an explanation of the POCS method. A description of both the convex and nonconvex constraints used in the superresolution problem is given, accompanied by results obtained from utilizing the different constraints to superresolve images.

4.2 INTRODUCTION TO SET THEORETIC ESTIMATION

Many image processing problems reduce to solving an optimization problem with or without constraints (e.g. minimum mean square error, maximum entropy, etc.). Different ways may be available to solve the problem, each producing a different solution, which may make the reliability of each solution questionable. The reliability of a solution should be based on all the knowledge available *a priori* and the observed data.

Every item of prior information reduces our ignorance of the problem to be solved [Combettes and Civanlar 1991].

In the past three decades, work has been done in finding a solution which satisfies all information available *a priori*. These are known as set theoretic estimation approaches, since each known property is associated with a set, and the intersection of all the sets represents the family of solutions, or feasibility set (family of objects that satisfy all the requirements). Any solution in the intersection set is a feasible solution since it is consistent with all the *a priori* information available [Katsaggelos 1989]. Set theoretic estimation can be defined as *an estimation framework in which consistency of a solution with the observed data and all a priori knowledge serves as the criterion of acceptability* [Combettes 1993].

In many cases the *a priori* information can be classified into three types:

- Information about the solution: this is knowledge available regarding properties of the result, such as signal intensity ranges, nonnegativity, etc.

- Information about the system: this is information regarding properties of the physical imaging system. This type of information is incorporated in the problem formulation, e.g. if the recorded data G is related to the object f (with Fourier transform F) through an operator $\mathbb{H}$, the transfer function of the imaging system, i.e. $G = \mathbb{H}F$, then any proposed solution $\hat{f}$ must satisfy $G = \mathbb{H}\hat{F}$, where $\hat{F}$ is the Fourier transform of $\hat{f}$.

- Information about external factors: this is information regarding unmeasurable external factors which affect the solution such as model uncertainty, noise, etc. This type of information is also incorporated within the problem formulation, e.g. if $G = \mathbb{H}F + N_o$ where N_o is the noise, the proposed solution $\hat{f}$ should also satisfy $G = \mathbb{H}\hat{F} + N_o$.

The algorithms which are based on the set theoretic estimation approach display best convergence when the property sets are closed and convex (explained below) and belong to the Hilbert space $\mathcal{H}$ [Hutson and Pym 1980].

4.3 CONVEXITY

The convexity property can be defined as:

Definition 4.1 *A subset C of a Hilbert space $\mathcal{H}$ is said to be convex if the line joining any two points in C is entirely contained in C, i.e. if z_1 and z_2 are points in C, then $\mu z_1 + (1-\mu)z_2$ must be contained in C for all $0 \leq \mu \leq 1$* [McCormick 1983].

This definition can be explained geometrically as stating that a set is convex if every single point on the line joining any two points in the set, belongs to the set. This property is illustrated in Fig. 4.1.

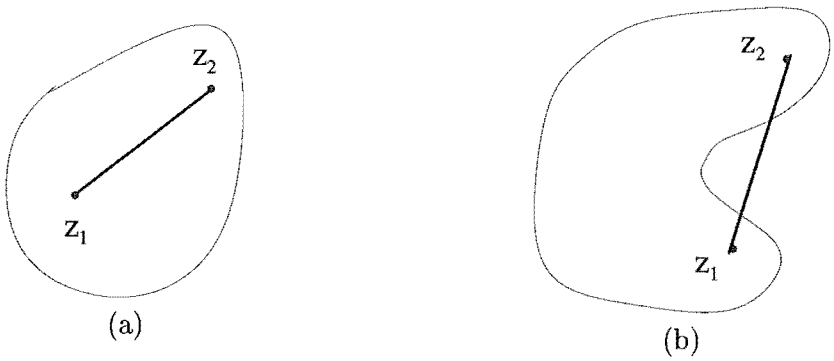


Figure 4.1 (a) Convex set; (b) nonconvex set.

It is also important to state some set operations that preserve convexity. Some of the common set operations are [Luenberger 1973]:

- If the set $\mathcal{C}$ is convex and β is a real number then the set $\beta\mathcal{C}$ is also convex, where $\beta\mathcal{C} = \{\beta z, z \in \mathcal{C}\}$.
  Proof: If $z_1, z_2 \in \mathcal{C}$, we have

$$\mu z_1 + (1-\mu)z_2 \quad \in \quad \mathcal{C},$$

  therefore

$$\mu\beta z_1 + (1-\mu)\beta z_2 \quad \in \quad \beta\mathcal{C}, \tag{4.1}$$

  which implies that $\beta\mathcal{C}$ is a convex set.

- If sets $\mathcal{C}_1$ and $\mathcal{C}_2$ are convex, then the set $\mathcal{C}_1 + \mathcal{C}_2$ is also convex, where $\mathcal{C}_1 + \mathcal{C}_2 = \{z_1 + z_3 : z_1 \in \mathcal{C}_1, z_3 \in \mathcal{C}_2\}$.
  Proof: If $z_1, z_2 \in \mathcal{C}_1$ and $z_3, z_4 \in \mathcal{C}_2$, where $\mathcal{C}_1$ and $\mathcal{C}_2$ are convex sets, we have

$$\mu z_1 + (1-\mu)z_2 \quad \in \quad \mathcal{C}_1 \tag{4.2}$$

  and

$$\mu z_3 + (1-\mu)z_4 \quad \in \quad \mathcal{C}_2, \tag{4.3}$$

  Adding Eqns. 4.2 and 4.3, we have

$$\mu(z_1 + z_3) + (1-\mu)(z_2 + z_4) \in \mathcal{C}_1 + \mathcal{C}_2. \tag{4.4}$$

Therefore the set $C_1 + C_2$ is a convex set.

- The intersection of any number of convex sets is also convex.
  Proof: Assume points z_1 and z_2 belong to the set C_0 which is the intersection of a finite number of convex sets C_i, i.e. $C_0 = \cap_{i=1}^{m} C_i$. Since z_1 and z_2 belong to the intersection C_0, then they belong to each set C_i. Since each C_i is convex, then $\mu z_1 + (1-\mu) z_2$ belongs to each set C_i and therefore belongs to the intersection C_0. Therefore C_0 is a convex set [McCormick 1983].

These properties are illustrated in Fig. 4.2.

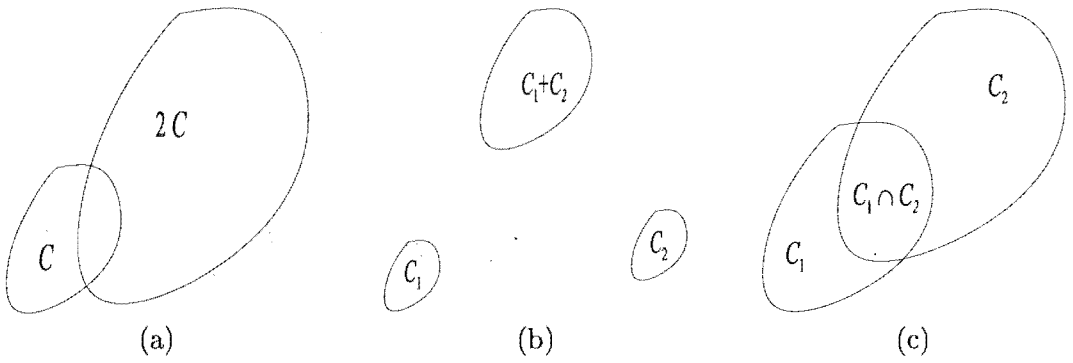


Figure 4.2 Graphical illustration of the properties of convex sets: (a) multiplication by a scalar; (b) addition; and (c) intersection.

4.4 CLOSURE

The other important property is closure, where a set is said to be closed if every point that is arbitrarily close to the set C belongs to C, i.e. C is closed if both $z_k \in C$ and the limit $z_k \to z$ imply $z \in C$ [Martos 1975, Luenberger 1984]. Fig. 4.3(a) illustrates an example of an open set $\{z : |z| < 1\}$ whereas Fig. 4.3(b) shows a closed set $\{z : |z| \leq 1\}$. Another example illustrating closure is: consider the end points z_1 and z_2 of a finite interval, then $(z_1, z_2) = \{z : z_1 < z < z_2\}$ is open whereas $[z_1, z_1] = \{z : z_1 \leq z \leq z_2\}$ is closed.

Figure 4.3 (a) An open set; (b) a closed set.

4.5 PROJECTION OPERATOR

As mentioned previously, every known property of the object f (provided the property meets certain criteria) will restrict it to lie in a closed convex set $\mathcal{C}_i$. Therefore, if m such properties are known then f is restricted to lie in the intersection of all the convex sets $\mathcal{C}_i, i = 1 \ldots m$, i.e. $f \in \mathcal{C}_0 = \cap_{i=1}^{m} \mathcal{C}_i$. The aim is to find a point in $\mathcal{C}_0$ and this is achieved by utilizing the projection operator $\mathcal{P}_i$ associated with each $\mathcal{C}_i$ [Yeh and Stark 1990]. The following theorem leads to the idea of the projection operator.

Theorem 4.1 *There exists a unique point $z_0 \in \mathcal{C}$, with $\mathcal{C}$ being a closed convex subset of $\mathcal{H}$, such that*

$$\min_{z \in \mathcal{C}} \|f - z\| = \|f - z_0\|, \tag{4.5}$$

where $f \in \mathcal{H}$.

Proof: Assume that

$$\min_{z \in \mathcal{C}} \|f - z\| = \delta = \|f - z_0\| = \|f - w_0\|,$$

where $z_0, w_0 \in \mathcal{C}$. Replacing z and w by $f - z_0$ and $f - w_0$ respectively in the following identity

$$\begin{aligned} \|\frac{z+w}{2}\|^2 + \|\frac{z-w}{2}\|^2 &= \frac{\|z\|^2 + \|w\|^2}{2} \\ \text{we have} \quad \|f - \frac{z_0+w_0}{2}\|^2 &= \delta^2 - \|\frac{z_0-w_0}{2}\|^2 \leq \delta^2. \end{aligned} \tag{4.6}$$

Since $\mathcal{C}$ is a closed convex set, therefore the midpoint $\frac{z_0+w_0}{2} \in \mathcal{C}$, (according to 4.1), so that

$$\|f - \frac{z_0+w_0}{2}\|^2 \geq \delta^2, \tag{4.7}$$

therefore $z_0 - w_0 = 0$, i.e. $z_0 = w_0$.

Definition 4.2 *For any $z \in \mathcal{H}$, the projection of z onto $\mathcal{C}$, $\mathcal{P}_{\mathcal{C}} z$, is the element in $\mathcal{C}$ closest to z, thus [Stark 1987]*

$$\|z - \mathcal{P}_{\mathcal{C}} z\| = \min_{w \in \mathcal{C}} \|z - w\|, \tag{4.8}$$

where $\mathcal{P}_C$ is the projection operator for $\mathcal{C}$.

4.6 THE METHOD OF POCS

As indicated in the introduction, in many medical and industrial applications where complete sets of data are impossible to collect, projection methods can be applied to the superresolution problem. The projection method is based on the idea that each constraint applied to an image f restricts it to lie in a corresponding set. If m properties are known, a solution is sought which lies at the intersection of all of the corresponding m sets. The superresolution problem using projection operators can be divided into two types:

- all constraints are convex;
- at least one constraint is nonconvex.

The first type can be treated with the POCS method, whereas the second type of problems can be treated by the method of generalized projections (MGP), explained in section 4.9.

POCS is a simple recursive algorithm which utilizes prior knowledge of the nature of the imaged object in the superresolution problem. Consider that m properties of the imaged object are known *a priori* and each corresponds to a convex set. There are thus m convex sets $\mathcal{C}_i$, $i = 1, \ldots, m$, and the intersection of the finite number of convex sets is also a convex set that contains f, i.e. $f \in \mathcal{C}_0 = \cap_{i=1}^{m} \mathcal{C}_i$ is also convex [Stark 1990]. If the projection operator $\mathcal{P}_0$ is known, then any $f \in \mathcal{H}$ can be projected onto $\mathcal{C}_0$, i.e. $\mathcal{P}_0 f \in \mathcal{C}_0$. However due to the complicated nature of $\mathcal{C}_0$, $\mathcal{P}_0$ is difficult to realize.

The approach of POCS is to recursively solve the problem by iteratively projecting the solution on to each $\mathcal{C}_i$ in turn using the corresponding $\mathcal{P}_i$, i.e.

$$\begin{aligned} z_n &= \mathcal{P}^n z, \qquad n = 0 \rightarrow \infty, \qquad \text{where} \\ \mathcal{P} &= \mathcal{P}_m \mathcal{P}_{m-1} \ldots \mathcal{P}_1. \end{aligned} \tag{4.9}$$

Fig. 4.4 is an example of the geometrical interpretation of Eqn. 4.9. In the simple example illustrated, $\mathcal{P} = \mathcal{P}_2 \mathcal{P}_1$, with $\mathcal{P}_1$ the projection operator onto a convex set $\mathcal{C}_1$, a closed sphere, and $\mathcal{P}_2$ the projection operator onto $\mathcal{C}_2$, a closed half space with boundary tangent to $\mathcal{C}_1$ at a point z^*. Therefore the intersection z^* is found by starting at an arbitrary point z then recursively generating the iterates z_n by utilizing Eqn. 4.9.

4.7 CONVEX CONSTRAINTS FOR THE SUPERRESOLUTION PROBLEM

In this section, those different constraints for the superresolution problem which correspond to convex sets, are discussed. Each constraint is defined and a proof of its convexity is given.

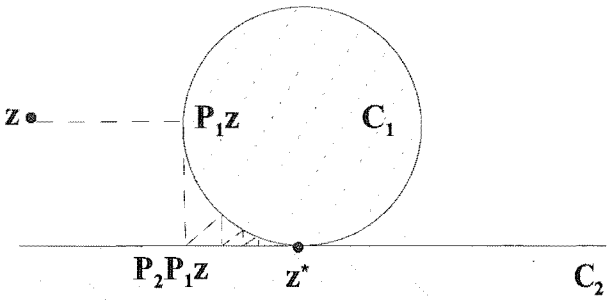


Figure 4.4 Example of geometrical representation of Eqn. 4.9.

4.7.1 The compact support constraint ($\mathcal{P}_1$)

$\mathcal{P}_1$ projects onto a set $\mathcal{C}_1$ all functions that vanish outside a prescribed region Γ. The constraint can be written as

$$\mathcal{P}_1 f = \begin{cases} f & (x,y) \in \Gamma \\ 0 & (x,y) \notin \Gamma. \end{cases} \tag{4.10}$$

To show convexity, assume $f_1, f_2 \in \mathcal{C}_1$ and $0 \leq \mu \leq 1$, then

$$\mu f_1 = \begin{cases} \mu f_1 & (x,y) \in \mu\Gamma \\ 0 & (x,y) \notin \mu\Gamma, \end{cases} \tag{4.11}$$

and

$$(1-\mu) f_2 = \begin{cases} (1-\mu) f_2 & (x,y) \in (1-\mu)\Gamma \\ 0 & (x,y) \notin (1-\mu)\Gamma. \end{cases} \tag{4.12}$$

Adding the two equations we have

$$\mu f_1 + (1-\mu) f_2 = \begin{cases} \mu f_1 + (1-\mu) f_2 & (x,y) \in (\mu\Gamma + (1-\mu)\Gamma) = \Gamma \\ 0 & (x,y) \notin (\mu\Gamma + (1-\mu)\Gamma) = \Gamma. \end{cases} \tag{4.13}$$

Therefore $\mu f_1 + (1-\mu) f_2 \in \mathcal{C}_1$, which proves its convexity according to Def. 4.1.

4.7.2 The frequency domain constraint ($\mathcal{P}_2$)

$\mathcal{P}_2$ projects onto a set $\mathcal{C}_2$ all functions whose Fourier transforms assume a prescribed value $G(k_x, k_y)$ over a region Ω in the Fourier plane. The constraint can be written as

$$\mathcal{P}_2 f = \begin{cases} \mathcal{F}^{-1} G(k_x, k_y) & (k_x, k_y) \in \Omega \\ \mathcal{F}^{-1} F(k_x, k_y) & (k_x, k_y) \notin \Omega, \end{cases} \tag{4.14}$$

where F is the Fourier transform of f. To show convexity, assume $f_1, f_2 \in \mathcal{C}_2$, thus

$$\mu\mathcal{P}_2 f_1 = \begin{cases} \mu\mathcal{F}^{-1}G(k_x,k_y) & (k_x,k_y) \in \Omega \\ \mu\mathcal{F}^{-1}F_1(k_x,k_y) & (k_x,k_y) \notin \Omega, \end{cases} \tag{4.15}$$

and

$$(1-\mu)\mathcal{P}_2 f_2 = \begin{cases} (1-\mu)\mathcal{F}^{-1}G(k_x,k_y) & (k_x,k_y) \in \Omega \\ (1-\mu)\mathcal{F}^{-1}F_2(k_x,k_y) & (k_x,k_y) \notin \Omega, \end{cases} \tag{4.16}$$

Adding Eqns. 4.15 and 4.16, we have

$$\mu\mathcal{P}_2 f_1 + (1-\mu)\mathcal{P}_2 f_2 = \begin{cases} \mathcal{F}^{-1}G(k_x,k_y) & (k_x,k_y) \in \Omega \\ \mu\mathcal{F}^{-1}F_1(k_x,k_y) + (1-\mu)\mathcal{F}^{-1}F_2(k_x,k_y) & (k_x,k_y) \notin \Omega, \end{cases} \tag{4.17}$$

Therefore $\mathcal{C}_2$ is a convex set.

4.7.3 The positivity constraint ($\mathcal{P}_3$)

$\mathcal{P}_3$ projects onto a set $\mathcal{C}_3$ all nonnegative functions. The constraint can be written as

$$\mathcal{P}_3 f = \begin{cases} f & f \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.18}$$

Assume $f_1, f_2 \in \mathcal{C}_3$, i.e. $f_1 \geq 0$ and $f_2 \geq 0$, therefore $\mu f_1 \geq 0$ and $(1-\mu)f_2 \geq 0$, for $0 \leq \mu \leq 1$. So $\mu f_1 + (1-\mu)f_2 \geq 0 \Rightarrow \mu f_1 + (1-\mu)f_2 \in \mathcal{C}_3$, implying that $\mathcal{C}_3$ is a convex set.

4.7.4 The energy constraint ($\mathcal{P}_4$)

$\mathcal{P}_4$ projects onto a set $\mathcal{C}_4$ all functions that satisfy

$$\|f\|^2 = \int\int_\Gamma |f(x,y)|^2 dxdy \leq \mathcal{E}_{lim}, \tag{4.19}$$

where $\mathcal{E}_{lim}$ is an energy limit and Γ is the compact support for f. Assuming $f_1, f_2 \in \mathcal{C}_4$ therefore

$$\|\mu f_1 + (1-\mu)f_2\| \leq \mu\|f_1\| + (1-\mu)\|f_2\| \leq \mu\sqrt{\mathcal{E}_{lim}} + (1-\mu)\sqrt{\mathcal{E}_{lim}} = \sqrt{\mathcal{E}}_{lim}. \tag{4.20}$$

Thus $\mathcal{C}_4$ is a convex set. The constraint can be written as

$$\mathcal{P}_4 f = \begin{cases} f & \mathcal{E} \leq \mathcal{E}_{lim} \\ \sqrt{\frac{\mathcal{E}_{lim}}{\mathcal{E}}} f & \mathcal{E} > \mathcal{E}_{lim}, \end{cases} \tag{4.21}$$

where $\mathcal{E} = \int\int_{\Gamma} |f(x,y)|^2 dxdy$.

4.7.5 The upper and lower bound constraint ($\mathcal{P}_5$)

$\mathcal{P}_5$ projects onto a set $\mathcal{C}_5$ all functions whose amplitudes lie within a closed interval $[a,b]$. The constraint can be written as

$$\mathcal{P}_5 f = \begin{cases} a & f < a \\ f & a \leq f \leq b \\ b & f > b. \end{cases} \tag{4.22}$$

To show convexity, assume $f_1, f_2 \in \mathcal{C}_5$, therefore

$$\mu f_1 + (1-\mu) f_2 \geq \mu a + (1-\mu) a = a \tag{4.23}$$

$$\mu f_1 + (1-\mu) f_2 \leq \mu b + (1-\mu) b = b, \tag{4.24}$$

therefore $\mathcal{C}_5$ is a convex set.

4.8 APPLICATION OF POCS

The POCS iteration is

$$\begin{aligned} f_{POCS}^{k+1} &= \mathcal{P} f_{POCS}^{k}, \\ \mathcal{P} &= \mathcal{P}_m \mathcal{P}_{m-1} \ldots \mathcal{P}_1. \end{aligned} \tag{4.25}$$

The G-P algorithm and its linear variations are special cases of the POCS method. Thus for the basic G-P algorithm

$$f_{G-P}^{k+1} = \mathcal{P}_2 \mathcal{P}_1 f_{G-P}^{k} \tag{4.26}$$

where $\mathcal{P}_1$ and $\mathcal{P}_2$ are the compact support and frequency domain constraints respectively, as given in sections 4.7.1 and 4.7.2.

Each of the following constraints was applied on both the bandlimited head and brain images (shown in Figs. 3.4(b) and 3.5(b) respectively), obtained from truncating the spectrum of the high resolution head and brain images (shown in Figs. 3.4(a) and 3.5(a) respectively), as discussed in section 3.7. Each constraint was studied by applying it on the recovered image at each iteration of the G-P algorithm. Since the G-P algorithm inherently applies both the compact support and the frequency domain constraints (i.e. constraints $\mathcal{P}_1$ and $\mathcal{P}_2$) and results applying the G-P algorithm have already been presented, results presented below are for constraints $\mathcal{P}_3$ to $\mathcal{P}_5$ only.

- The compact support constraint ($\mathcal{P}_1$) sets all pixels of the image f^k outside the

support Γ to zero. The constraint can be written as

$$\mathcal{P}_1 f^k = \begin{cases} f^k & (x,y) \in \Gamma \\ 0 & (x,y) \notin \Gamma. \end{cases} \tag{4.27}$$

- The frequency domain constraint ($\mathcal{P}_2$) corrects the spectrum within the bandlimit (Ω_x, Ω_y) to the measured $G(k_x, k_y)$. The constraint can be written as

$$\mathcal{P}_2 f^k = \begin{cases} \mathcal{F}^{-1} G(k_x, k_y) & |k_x| \leq \Omega_x, |k_y| \leq \Omega_y \\ \mathcal{F}^{-1} F^k & |k_x| > \Omega_x, |k_y| > \Omega_y. \end{cases} \tag{4.28}$$

- The positivity constraint ($\mathcal{P}_3$) sets all pixels of f^k with negative values to zero. The constraint can be written as

$$\mathcal{P}_3 f^k = \begin{cases} f^k & f^k \geq 0 \\ 0 & otherwise. \end{cases} \tag{4.29}$$

The use of the positivity constraint is justified since the three quantities which are imaged in MRI, T_1, T_2 and proton density, are all inherently nonnegative. However due to some external factors such as noise and to the fact that not all the Fourier domain is measured, negative values may result in the recovered image which are clearly in error. Fig. 4.5(a) shows the recovered image (at iteration 6) obtained by applying the positivity constraint on the bandlimited brain image and Fig. 4.5(b) shows the corresponding ISNR curve. There is a slight improvement in the ISNR curve compared to that obtained by applying the G-P algorithm without the positivity constraint, but there is limited visual improvement in the recovered image compared to the bandlimited, due to the fact that there are very few negative pixels (which are corrected by this constraint) within the support region as shown in Fig. 4.6.

- The energy constraint ($\mathcal{P}_4$) restricts the image f^k to satisfy

$$\int\int_{\Gamma} |f^k(x,y)|^2 dxdy = \mathcal{E}^k \leq \mathcal{E} + \delta\mathcal{P}_4 = \mathcal{E}_{lim}, \tag{4.30}$$

where $\mathcal{E}^k$ is the energy of the recovered (superresolved) image, $\mathcal{E}$ is the energy of the bandlimited image, $\delta\mathcal{P}_4$ is the incremental energy associated with the recovery of the additional spatial frequencies and $\mathcal{E}_{lim}$ is an energy limit. The constraint can be written as

$$\mathcal{P}_4 f^k = \begin{cases} f^k & \mathcal{E} \leq \mathcal{E}_{lim} \\ \sqrt{\frac{\mathcal{E}_{lim}}{\mathcal{E}}} f^k & \mathcal{E} > \mathcal{E}_{lim}. \end{cases} \tag{4.31}$$

Both the energy and positivity constraint were applied on the bandlimited head

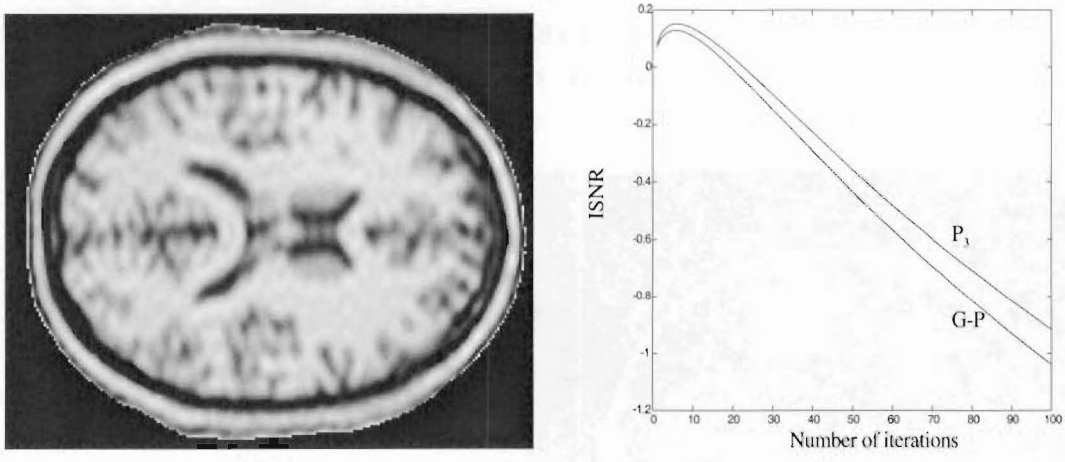


Figure 4.5 Positivity constraint: (a) recovered brain image; (b) ISNR curve.

Figure 4.6 Pixels of the bandlimited brain image with negative values.

image. The energy constraint was applied such that the energy of the recovered image in each iteration was within a deviation $\delta\mathcal{P}_4$ from the energy of the bandlimited head image, where $\delta\mathcal{P}_4$ is expressed as a percentage. Fig. 4.7(a) shows the recovered image (at iteration 1000) for $\delta\mathcal{P}_4 = 1\%$. Again there was limited visual improvement compared to the bandlimited head image, however the ISNR plot shown in Fig. 4.7(b) shows that the algorithm is more stable and no longer so sensitive to the number of iterations performed.

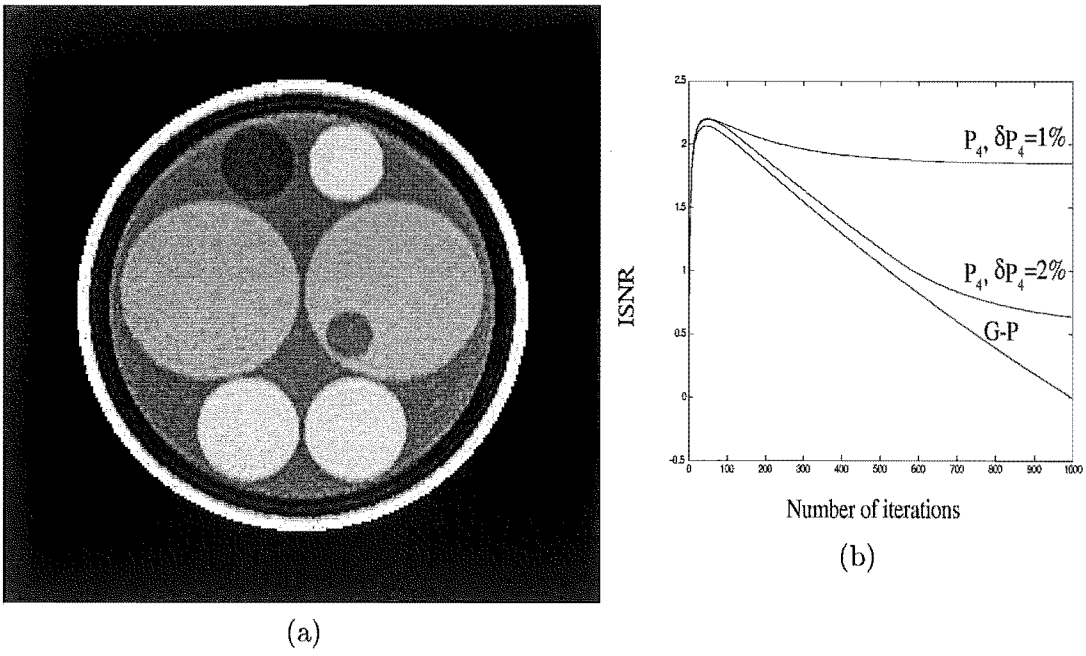


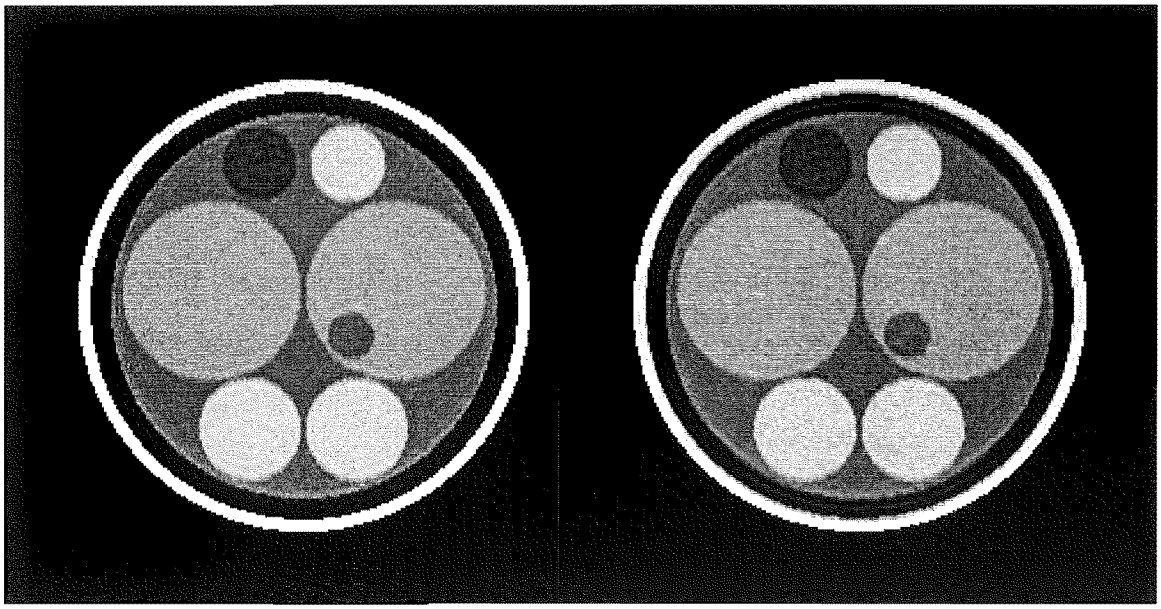
Figure 4.7 Energy constraint: (a) recovered head image with $\delta\mathcal{P}_4 = 1\%$; (b) ISNR curve.

- The lower and upper bound constraint ($\mathcal{P}_5$) limits pixel values to physically meaningful upper and lower bounds. The constraint can be written as

$$\mathcal{P}_5 f^k = \begin{cases} a & f^k < a \\ f^k & a \leq f^k \leq b \\ b & f^k > b, \end{cases} \tag{4.32}$$

where a and b are the lower and upper bounds respectively. The purpose of using this constraint on realistic MR images is that physically unrealistic pixel values, arising because of noise, errors in reconstruction or limitations of the measurement process, are corrected to more meaningful values. The main drawback of this method is that exact lower and upper bounds are not known in most practical applications, therefore only approximations to these bounds can be used. The $\mathcal{P}_5$ operator was applied on the bandlimited head model. The exact lower and upper bounds of the high resolution head phantom are $[0.1, 1]$, where the lower bound was considered to be the minimum of the high resolution image within

the support constraint only. The bounds that were used in the recovery process were $a - \delta\mathcal{P}_5$ and $b + \delta\mathcal{P}_5$ where $[a, b] = [0.1, 1]$ and $\delta\mathcal{P}_5 = -0.2, 0, 0.1$. For both the cases when $\delta\mathcal{P}_5 = 0, 0.1$, there is limited visual improvement in the recovered images as shown in Figs. 4.8(a) and (b), even though there is a considerable improvement in the ISNR curve when $\delta\mathcal{P}_5 = 0$. The reason behind this is that there are not many pixels (within the support region) in the bandlimited image that are outside the bounds $[0.1, 1]$. Fig. 4.9(a) shows the recovered image for the case $\delta\mathcal{P}_5 = -0.2$ (underestimation of bounds), the recovered image appears worse with many pixels set to incorrect values.



(a) (b)

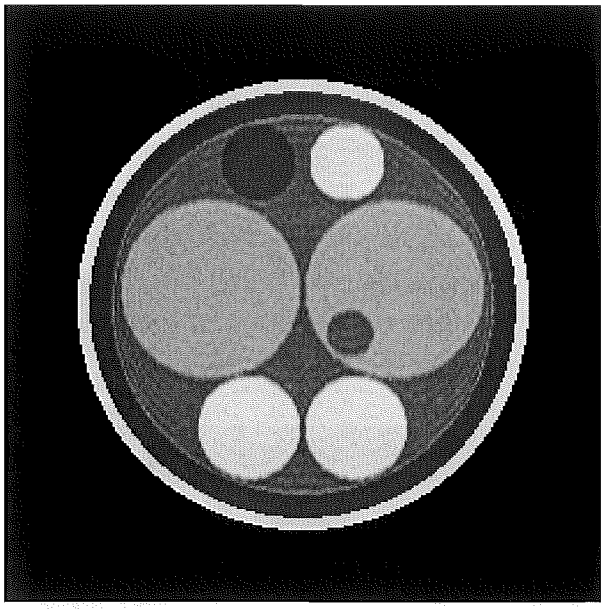
Figure 4.8 Recovered image using lower and upper bound constraints with (a) $\delta\mathcal{P}_5 = 0$; (b) $\delta\mathcal{P}_5 = 0.1$.

4.9 THE METHOD OF GENERALIZED PROJECTIONS (MGP)

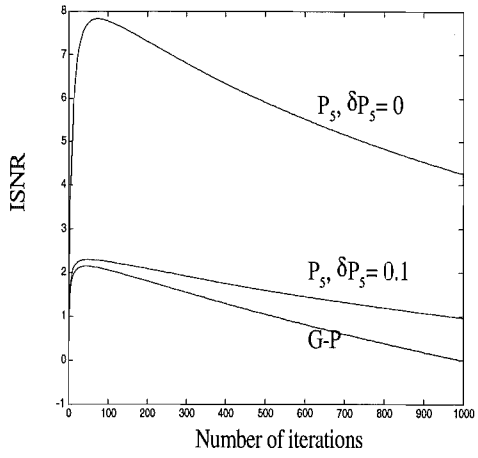
The MGP is an extension of the POCS method to cases where the *a priori* constraints are nonconvex [Katsaggelos 1991]. An example of a nonconvex set is quantization of a signal. The projection operator for the MGP method can be defined as

$$\|g - x\| = \min_{\text{all } y \in \mathcal{C}} \|y - x\|, \tag{4.33}$$

where $g = \mathcal{P}_{\mathcal{C}} x$ is the projection of x onto $\mathcal{C}$. If $\mathcal{C}$ was a convex set, the solution would have been unique, but when $\mathcal{C}$ is nonconvex there might be a set of points satisfying Eqn. 4.33 and convergence is not guaranteed. Another problem that can arise with the MGP algorithm is that it might get trapped or end up in a tunnel due to the



(a)



(b)

Figure 4.9 Lower and upper bound constraint: (a) recovered head image, with $\delta\mathcal{P}_5 = -0.2$; (b) ISNR curve.

nonconvexity of one of the constraints. A trap is a fixed point which fails to satisfy one or more of the constraints but satisfies

$$f^{k+1} = \mathcal{P}_m\mathcal{P}_{m-1}\ldots\mathcal{P}_1 f^k = f^k, \tag{4.34}$$

whereas a point f_k is said to be in a tunnel if the change in f_k from one iteration to the next in Eqn. 4.34 is negligible. It is important to note that the tunnel situation might also arise for the convex case, i.e., when all the sets are convex. The trap and tunnel situations are illustrated in Fig. 4.10(a) and (b) respectively. One simple way of getting out of a trap is to modify the initialization and repeat the procedure.

Since only limited superresolution was obtained from the application of convex constraints, discussed in section 4.8, nonconvex constraints were studied in the attempt to improve resolution of the recovered images. The nonconvex constraints studied are discussed below.

4.9.1 Single level quantization ($\mathcal{P}_6$)

This projection-based operator sets pixels within a certain deviation of a single *level* (the most frequently occurring) in the image to that *level*. The constraint can be written as

$$\mathcal{P}_6 f^k = \begin{cases} level & level - \delta\mathcal{P}_6 \leq f^k \leq level + \delta\mathcal{P}_6 \\ f^k & otherwise, \end{cases} \tag{4.35}$$

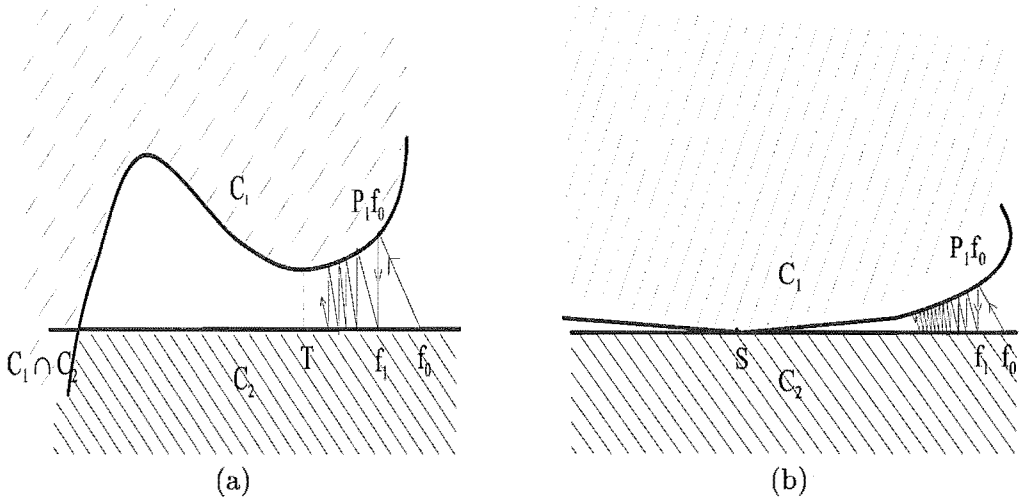


Figure 4.10 Examples of a trap and a tunnel for an algorithm of the form $f^{k+1} = \mathcal{P}_2\mathcal{P}_1 f^k$: (a) algorithm will converge to a trap point T, whereas the true solution is in the intersection of C_1 and C_2; (b) the algorithm will iterate slowly along the tunnel towards the solution S.

where $\delta\mathcal{P}_6$ is the deviation from the *level*. This projection-based operator is different from the conventional quantization as shown in Fig. 4.11(a) where $\eta_i = \frac{level_{i+1}+level_i}{2}$, $level_1 < level_2 < level_3$ and pixels with values between η_i and η_{i+1} are set to $level_{i+1}$. In this case every pixel is set to belong to a nonconvex set resulting from a nonconvex projection operation. This nonconvex projection operator is also nonunique since pixels with values η_i could be set to either $level_i$ or $level_{i+1}$. This situation is different from the single level quantization illustrated in Fig. 4.11(b), where firstly we consider a single level only (the most frequently occurring) where say $level_2 = level$ and deviation from the level is restricted such that pixels with values less than $level - \delta\mathcal{P}_6$ or greater than $level + \delta\mathcal{P}_6$ are left unchanged. For our practical algorithms we found better results with the single level quantization, since pixels for example with values between η_1 and $level - \delta\mathcal{P}_6$ have a higher uncertainty of having their value=$level$ than those with values between $level - \delta\mathcal{P}_6$ and $level + \delta\mathcal{P}_6$.

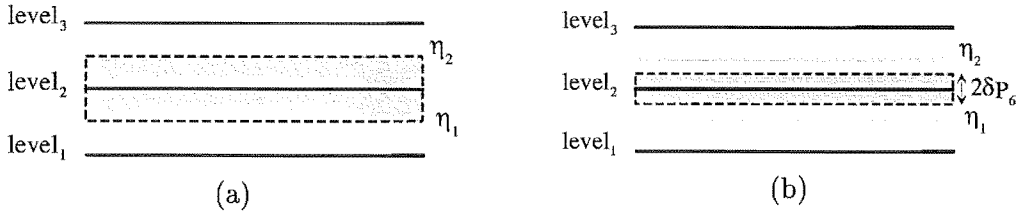
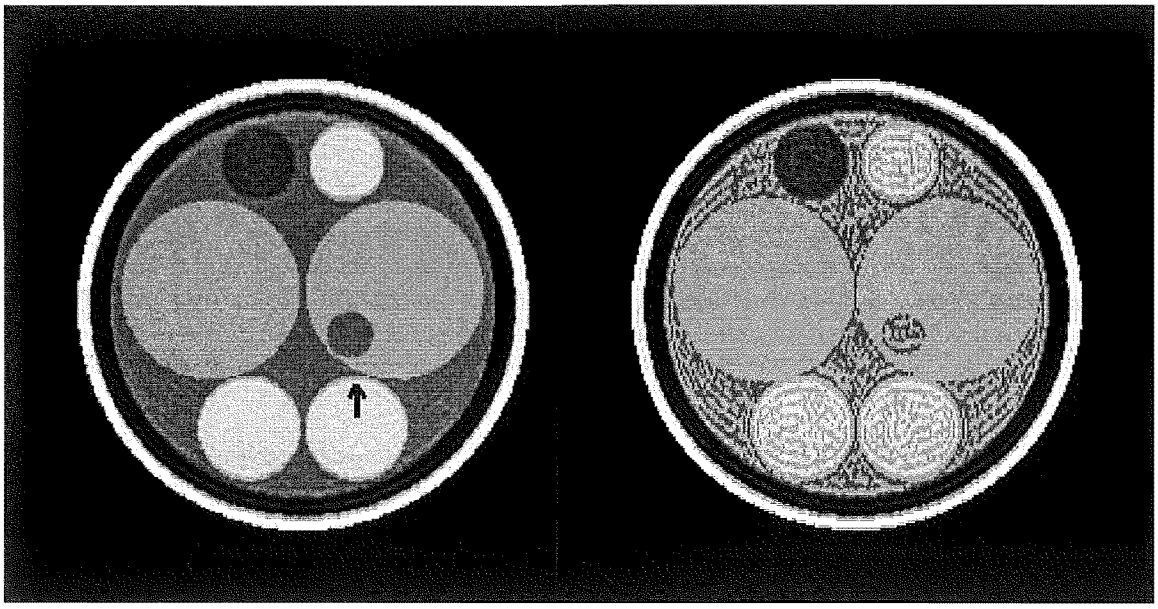


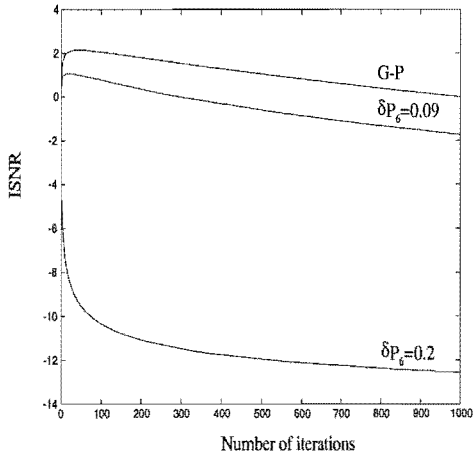
Figure 4.11 Illustration of: (a) conventional quantization; (b) single level quantization.

The $\mathcal{P}_6$ operator was applied on the bandlimited head image, with $level = 0.6$ (the level occurring most frequently in the original head image), and $\delta\mathcal{P}_6 = 0.09$ and 0.2, as shown in Fig. 4.12, where the recovered images were all obtained at the iteration with the maximum ISNR value. In this case the ISNR curves utilizing the $\mathcal{P}_6$ operator

appear worse compared to the ISNR curve obtained from the G-P algorithm, however visually the recovered image for the case $\delta\mathcal{P}_6 = 0.09$ appears much sharper and the edges are much more defined, however there are some pixel values (indicated by the arrow) incorrectly set to *level* $= 0.6$, and as $\delta\mathcal{P}_6$ increases to 0.2, so does the number of pixels whose values are incorrectly set to 0.6 and the image is degraded.



(a) (b)



(c)

Figure 4.12 Recovered head image using single level quantization with level=0.6 and (a) $\delta\mathcal{P}_6 = 0.09$, (b) $\delta\mathcal{P}_6 = 0.2$, (c) ISNR graph.

Figs. 4.13(a) and (b) show the histogram of the high resolution head image and the histogram of the bandlimited image respectively. There is significant difference between the two histograms. The aim of the $\mathcal{P}_6$ operator is to make the histogram of the recovered image approach that of the high resolution image. Fig. 4.13(c) shows the histogram of the recovered image using $\mathcal{P}_6$, with *level* $= 0.6, \delta\mathcal{P}_6 = 0.09$, where

all pixels with values between 0.51 and 0.69 were set to 0.6. The improvement in the recovered image is portrayed as the smaller difference between the histograms of the high resolution image and the recovered image compared to the difference between the histograms of the high resolution image and the bandlimited image. When $\delta\mathcal{P}_6$ was increased to 0.2, the recovered image shown in Fig. 4.12(b) appears much worse, since many pixels between levels 0.4 and 0.8 were incorrectly set to 0.6 as shown in the corresponding histogram in Fig. 4.13(d).

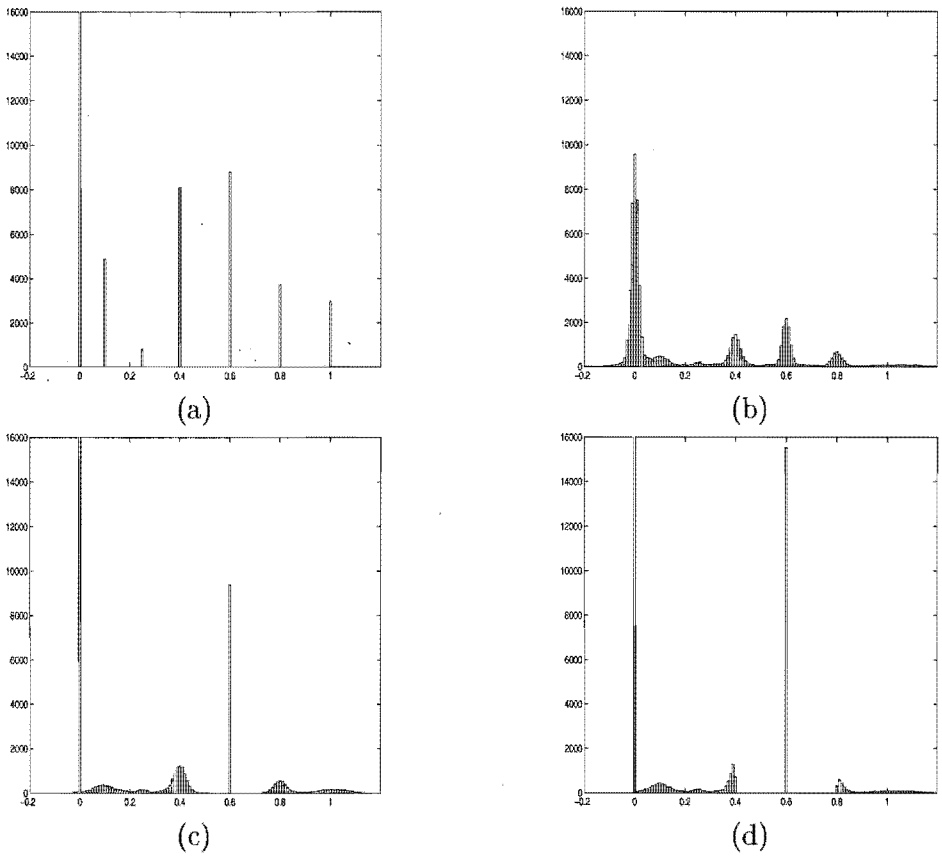


Figure 4.13 Histograms: (a) high resolution head model; (b) bandlimited noisy head model; (c) recovered image using $\mathcal{P}_6$, $level = 0.6, \delta\mathcal{P}_6 = 0.09$; (d) recovered image using $\mathcal{P}_6$, $level = 0.6, \delta\mathcal{P}_6 = 0.2$.

The single level quantization can be extended to multiple levels. This is achieved by modifying the constraint defined in Eqn. 4.35 to

$$\mathcal{P}_6 f^k = \begin{cases} level_i & level_i - \delta\mathcal{P}_6 \leq f^k \leq level_i + \delta\mathcal{P}_6 \\ f^k & otherwise, \end{cases} \tag{4.36}$$

where $level_i$ represent the different $level$ values and Eqn. 4.36 is applied for each $level$ within each iteration of the G-P algorithm, and will be referred to as multiple level quantization. The multiple level quantization was applied on the head data with $level_i = 0.1, 0.25, 0.4, 0.6$ and 0.8, and with $\delta\mathcal{P}_6 = 0.09$, resulting in a relatively sharp

recovered image, shown in Fig. 4.14.

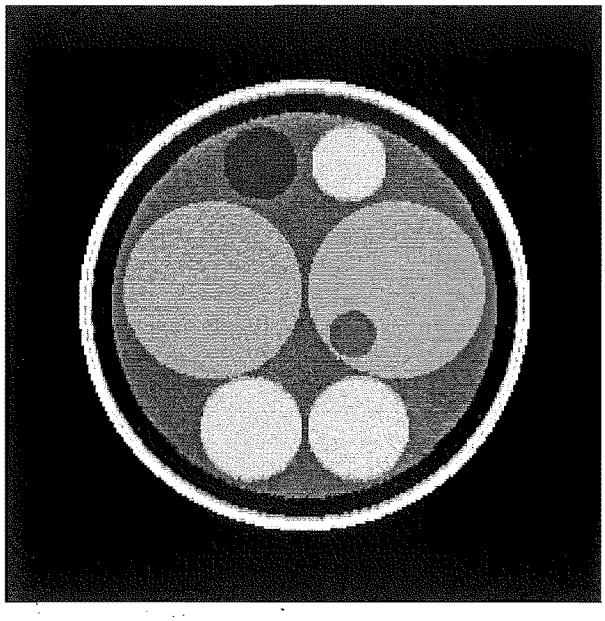


Figure 4.14 Recovered head image using multiple level quantization with $level_i = 0.1, 0.25, 0.4, 0.6$ and 0.8; $\delta\mathcal{P}_6 = 0.09$.

An alternative method that imposes a number of different levels on the solution is presented in section 8.6. The single level quantization constraint is a harsh constraint and is applied on each pixel independently without taking into consideration any relationship that could exist between the pixels. The next constraint was designed to ameliorate the above drawback.

4.9.2 Neighbourhood based quantization ($\mathcal{P}_7$)

This projection-based operator sets any pixel within a certain deviation of a $level_i$ to that $level_i$, if within a certain neighbourhood of the pixel considered (e.g. (3×3) as shown in Fig. 4.15(a)) the majority of pixels (i.e. five pixels or more) are within a certain deviation of that $level_i$. The constraint can be written as

$$\mathcal{P}_7 f^k(x,y) = \begin{cases} level_i & \Xi(x,y) = 1 \\ f^k(x,y) & otherwise, \end{cases} \tag{4.37}$$

where $\Xi(x,y) = 1$ if the majority of the 3×3 neighbourhood pixel values, centered on (x,y), lie within the range $level_i \pm \delta\mathcal{P}_7$, and $\Xi(x,y) = 0$, otherwise [Bones *et al.* 1998].

The $\mathcal{P}_7$ operator was also applied on the bandlimited head data, with $\delta\mathcal{P}_7 = 0.09$. The constraint was applied for 5 different levels simultaneously (where $level_i = 0.1, 0.25, 0.4, 0.6$ and 0.8). Fig. 4.15(b) shows the recovered image (at iteration 96) using operator $\mathcal{P}_7$, and it appears much sharper and more defined compared to the bandlimited head image and the recovered image obtained from applying the multiple

level quantization, shown in Fig. 4.14. This result is also expressed in the higher ISNR curve obtained from applying the $\mathcal{P}_7$ operator compared to the ISNR curve of the G-P algorithm as shown in Fig. 4.15(c). The $\mathcal{P}_7$ operator was also applied on the bandlimited brain image, with $\delta\mathcal{P}_7 = 10$. The recovered image is shown in Fig. 4.15(d). Even though the application of the operator does not provide extra information, the edges of the recovered image are sharper and the image itself is visually preferred to the bandlimited image. The $\mathcal{P}_7$ operator attempts to utilize the prior expectation that substantial areas of near homogeneous tissue exist in the slice being imaged. The main

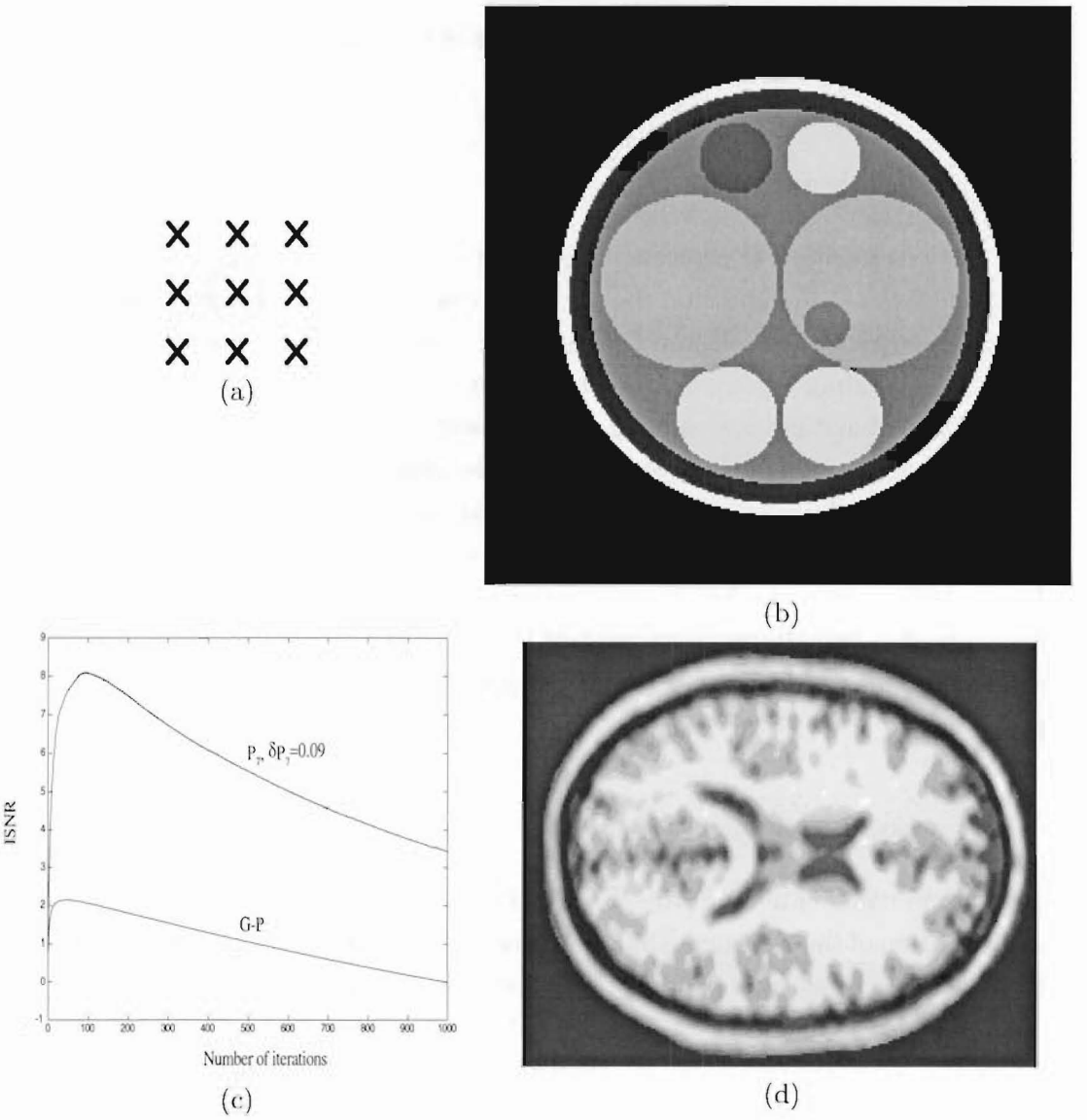


Figure 4.15 (a) 3×3 neighbourhood; neighbourhood based quantization: (b) recovered head image with $\delta\mathcal{P}_7 = 0.09$; (c) ISNR curve; (d) recovered brain image with $\delta\mathcal{P}_7 = 10$.

disadvantage of this method is that pixels lying between two very close but different tissues might be set to the values of either tissue, as shown in Fig. 4.15(b). It is

important to note that for noisy data the solution might diverge. Regularization, explained in chapter 5, can be used to reduce the likelihood of this happening.

4.10 COMBINATION OF PROJECTION OPERATORS

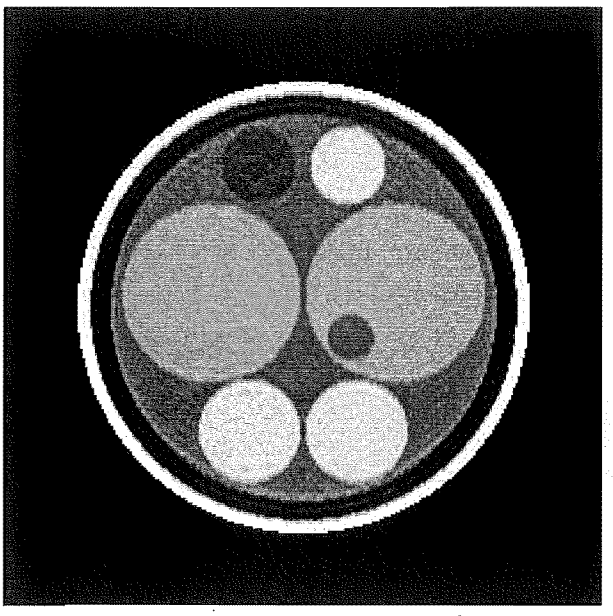
In this section, combinations of different projection operators are studied. The idea behind this is to benefit from the advantage of each operator. For example some operators result in an ISNR curve with a higher peak, whereas others results in a levelled off ISNR curve, so by combining these operators the aim is to obtain a levelled off ISNR curve with a peak higher than most of the peaks obtained by applying each constraint alone.

4.10.1 Combination A ($\tilde{\mathcal{P}}_A$)

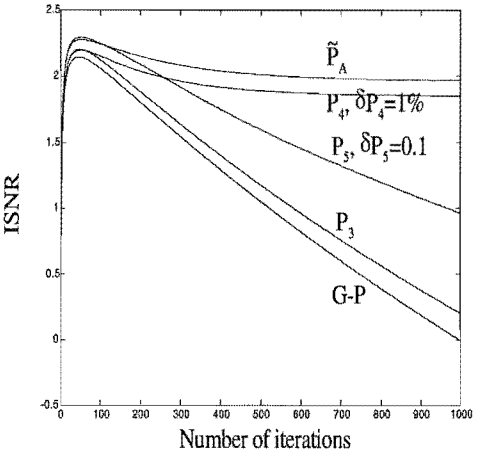
The first combination of projection operators studied is $\tilde{\mathcal{P}}_A = \mathcal{P}_5\mathcal{P}_4\mathcal{P}_3$, where in each iteration of the G-P algorithm, the positivity constraint is applied first followed by the energy constraint then followed by the upper and lower bound constraint. As before, the three operators are applied in addition to $\mathcal{P}_1$ and $\mathcal{P}_2$ of the basic G-P algorithm. This $\tilde{\mathcal{P}}_A$ combination was applied on the bandlimited head image with $\delta\mathcal{P}_4 = 1\%$, $\delta\mathcal{P}_5 = 0.1$. Since each of these constraints when applied independently did not improve the image visually, similarly there is no visual improvement in the recovered image obtained from combining these constraints as shown in Fig. 4.16(a) (at iteration 1000). Fig. 4.16(b) shows the ISNR curve obtained using $\tilde{\mathcal{P}}_A$. The ISNR curve seems to have a peak a little higher than most of the ISNR curves obtained from applying the constraints independently and it is no longer as sensitive to the number of iterations performed.

4.10.2 Combination B ($\tilde{\mathcal{P}}_B$)

The second combination of projection operators studied is $\tilde{\mathcal{P}}_B = \mathcal{P}_6\mathcal{P}_5\mathcal{P}_4\mathcal{P}_3$, where in each iteration of the G-P algorithm, the convex $\tilde{\mathcal{P}}_A$ constraint is applied first followed by the nonconvex single level quantization constraint. The $\tilde{\mathcal{P}}_B$ operator was applied on the bandlimited head image with $\delta\mathcal{P}_4 = 1\%$, $\delta\mathcal{P}_5 = 0.1$, level=0.6 and $\delta\mathcal{P}_6 = 0.09$. Fig. 4.17 shows the recovered image (at iteration 1000) and the corresponding ISNR curve. In this case the ISNR obtained from applying the $\tilde{\mathcal{P}}_B$ operator is a little higher and levelled off compared to the ISNR curve obtained from applying $\mathcal{P}_6$ independently, however the recovered image (at iteration 1000) obtained from applying the $\tilde{\mathcal{P}}_B$ operator appears worse with more ringing artifact and less sharp edges compared to the recovered image (at iteration with maximum ISNR) obtained from applying $\mathcal{P}_6$ independently (shown in Fig. 4.12(a)).

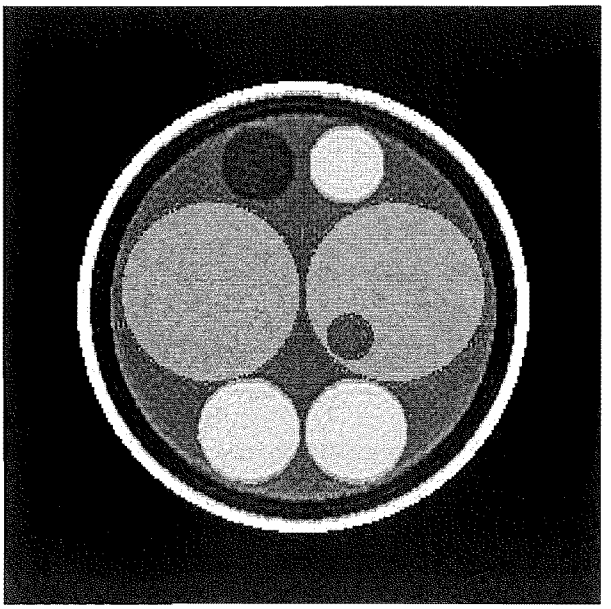


(a)

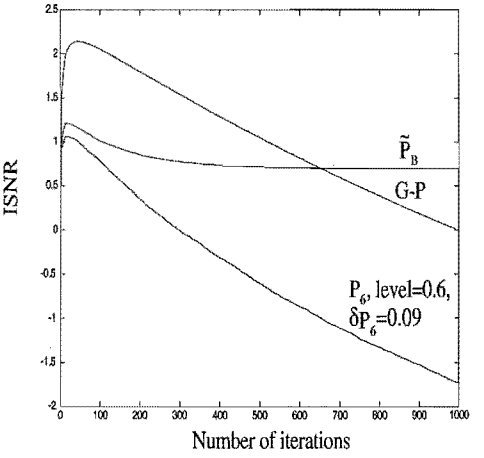


(b)

Figure 4.16 (a) Recovered head image using the combination projection operators $\tilde{\mathcal{P}}_A$, with $\delta\mathcal{P}_4 = 1\%$, and $\delta\mathcal{P}_5 = 0.1$; (b) ISNR curve.



(a)

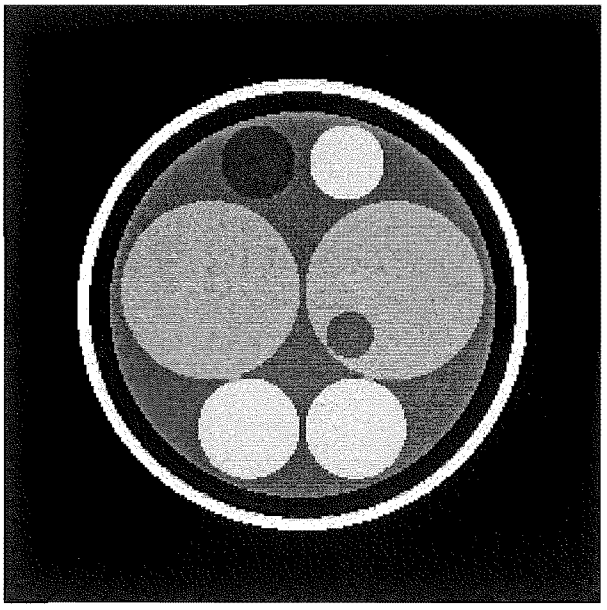


(b)

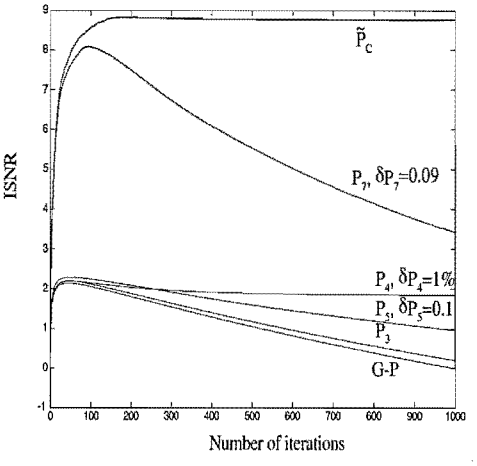
Figure 4.17 (a) Recovered head image using the combination projection operators $\tilde{\mathcal{P}}_B$ with $\delta\mathcal{P}_4 = 1\%$, $\delta\mathcal{P}_5 = 0.1$, level=0.6 and $\delta\mathcal{P}_6 = 0.09$; (c) ISNR curve.

4.10.3 Combination C ($\tilde{\mathcal{P}}_C$)

Another combination of projection operators studied is $\tilde{\mathcal{P}}_C = \mathcal{P}_7\mathcal{P}_5\mathcal{P}_4\mathcal{P}_3$, where in this case the operator $\tilde{\mathcal{P}}_C$ projects onto both convex and nonconvex sets. Fig. 4.18(a) shows the recovered image (at iteration 1000) obtained from applying the $\tilde{\mathcal{P}}_C$ operator on the bandlimited head image and Fig. 4.18(b) shows the corresponding ISNR curve in comparison to the ISNR curves obtained from applying the G-P algorithm and the $\mathcal{P}_7$ operator. In this case the ISNR curve obtained from applying $\tilde{\mathcal{P}}_C$ has a higher peak from that obtained from applying $\mathcal{P}_7$ alone and is no longer sensitive to the number of iterations performed.



(a)



(b)

Figure 4.18 (a) Recovered head image using the combination projection operators $\tilde{\mathcal{P}}_C$ with $\delta\mathcal{P}_4 = 1\%$, $\delta\mathcal{P}_5 = 0.1$ and $\delta\mathcal{P}_7 = 0.09$; (b) ISNR curve.

4.11 CONCLUSION

Different constraints, corresponding to both convex and nonconvex sets, were studied and applied to both the head and brain bandlimited data. There was only a slight improvement in the recovered brain image when imposing the positivity constraint due to the absence of any negative values in the bandlimited image. However when applying the G-P algorithm, some negative values were introduced and the positivity constraint corrected them by setting them to zero at each iteration. The upper and lower bound constraint was found to improve the performance only when knowledge of exact bounds is available which is not usually the case in most practical applications. Constraints corresponding to nonconvex sets were also studied, and the neighbourhood

based quantization constraint seems to provide the best result due to the fact that it takes into consideration the relationship that might exist between neighbouring pixels.

In spite of the popularity of POCS in the signal processing field, it suffers from three major drawbacks. The first is its serial algorithmic structure which means that only one set is activated at any point in time. Therefore it is impractical to implement on a system with parallel processors [Combettes 1994]. The second disadvantage of POCS is its slow convergence, which can be slightly accelerated by relaxing the projections at each iteration [Combettes 1994]. The third drawback of POCS arises if two or more constraints are incompatible. This might cause the intersection of the property sets to be an empty set, so that in this case the best solution that POCS can provide can belong to only one set [Combettes and Puh 1993].

Chapter 5

REGULARIZATION

5.1 INTRODUCTION TO REGULARIZATION

The concept of well-posed problems was first introduced by Hadamard early in the 20^{th} century when studying the types of boundary conditions that are most natural for various types of differential equations. For many years it was believed that physical situations resulted in only well-posed problems and that ill-posed problems were just of academic interest. However later it was found that many physical situations result in ill-posed problems [Baltes 1980]. Since most problems attempt to solve $\mathcal{A}u = b$, where $u \in \mathcal{U}$, $b \in \mathcal{B}$ and $\mathcal{A}$ is a bounded linear operator, this equation is useful in explaining the meaning of a well-posed problem. The problem of solving for u given b is said to be well-posed if

- for every element $b \in \mathcal{B}$ there exists a solution $u \in \mathcal{U}$;
- the solution is unique; and
- the problem is stable (i.e. the solution continuously depends on the data).

If any of these conditions are not met, the problem is said to be ill-posed [Katsaggelos 1991]. The first condition can always be met by increasing the solution space, and if more than one solution is obtained, then additional information can be incorporated into the problem to obtain a unique solution. The third condition is the most important one, since it is practically impossible to find a solution for an unstable problem, where the solution does not continuously depend on the data [Kirsch 1996]. Differentiation of a function $b(t)$ that is only known approximately is an example of an ill-posed problem which requires regularization to solve. To illustrate this, let $u_1(t)$ be the first derivative of $b_1(t)$. Assume an approximation to $b_1(t)$, $b_2(t) = b_1(t) + C\sin(\omega t)$, is available. The maximum absolute difference between $b_1(t)$ and $b_2(t)$ is $|C|$, whereas the maximum absolute difference between $u_1(t)$ and $u_2(t) = \frac{d}{dt}b_2(t)$ is $|C\omega|$, which can be made arbitrarily large by choosing large values of ω.

In the imaging process, images are usually degraded due to either imperfections of the imaging systems or to physical limitations of the imaging process (often these

differ between different imaging techniques). The degradation changes or removes some of the spatial frequency content of the object being imaged [Saito 1990]. In addition to being degraded, the measured data suffers from the presence of noise. Therefore, only approximations to the true images are recorded. For the continuous 2-D case, the superresolution problem (corresponding to Eqn. 3.4 for the discrete case) can be written as

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathfrak{h}(x,y,\xi,\eta) f(\xi,\eta) d\xi d\eta, \tag{5.1}$$

where g is the image, $\mathfrak{h} = \mathcal{F}^{-1}(\mathbb{H})$, where $\mathbb{H}$ is the transfer function of the image formation system and f is the imaged object. The superresolution problem can be proven to be ill-posed by making use of the Riemann-Lebesgue lemma, which states that if $\mathfrak{h}(\xi,\eta)$ is an integrable function then it can be shown that [Apostol 1958]

$$\lim_{\beta\to\infty}\lim_{\alpha\to\infty}\int_a^b\int_a^b \mathfrak{h}(\xi,\eta)\sin(\alpha\xi)\sin(\beta\eta)d\xi d\eta = 0, \tag{5.2}$$

which implies that for a small value $\epsilon_1 > 0$, there exists a value Q such that

$$\int_a^b\int_a^b \mathfrak{h}(x,y,\xi,\eta)\sin(\alpha\xi)\sin(\beta\eta)d\xi d\eta < \epsilon_1, \tag{5.3}$$

whenever $\alpha,\beta \geq Q$. Therefore,

$$\begin{aligned}\int_a^b\int_a^b \mathfrak{h}(x,y,\xi,\eta)[f(\xi,\eta) + \sin(\alpha\xi)\sin(\beta\eta)]d\xi d\eta &= \\ \int_a^b\int_a^b \mathfrak{h}(x,y,\xi,\eta) f(\xi,\eta) d\xi d\eta + \epsilon &= g(x,y) + \epsilon,\end{aligned} \tag{5.4}$$

for all values of $\alpha,\beta \geq Q$, where $\mid \epsilon \mid < \epsilon_1$. Therefore, any small perturbation in g cannot be distinguished from a non-trivial additional component of a sinusoid with frequency $\alpha,\beta \geq Q$ in the original image [Andrews and Hunt 1977]. The method of regularization is then used to overcome the ill-posed nature of the superresolution problem.

This chapter starts with a general overview of the regularization method and its uses. Different regularization methods are studied, such as the Tikhonov-Miller regularization method in section 5.3, and the model-based regularization in section 5.4. Depending on the choice of the regularizing operator, either maximum smoothness or edge-preserving regularization may be achieved. Each of the above methods are discussed and the corresponding results are given.

5.2 THE REGULARIZATION METHOD

In this work, most of the superresolving techniques are iterative due to the following advantages [Katsaggelos *et al.* 1991]:

- iterative techniques do not need to determine the inverse of the transfer function of the imaging system which allows large images to be superresolved,
- prior knowledge of properties of the solution can be easily incorporated within the method,
- the progress of the solution can be monitored.

However one of the problems of the iterative methods discussed up to this point (i.e. the G-P and POCS algorithms) is that the performance (ISNR) increases to a maximum and then starts to decrease as the number of iterations increases, due to the non-convergence of the algorithm in the presence of noise. The aim of using regularization is then to level off the ISNR graph so that it is no longer sensitive to the number of iterations once the maximum is achieved.

The superresolution problem for an MR image can be written as $Hf = G$, where H is the operator which models the transfer function of the imaging system, G is the collected data (in the spatial frequency domain), and f is the imaged object. Suppose the exact solution of $Hf = G$ is f_T which belongs to set Θ, therefore $H\Theta$ is the mapping of set Θ by the operator H. The exact solution to $Hf = G$ is obtained only when G belongs to the set $\mathcal{G} = H\Theta$ (i.e. $G = G_T = Hf_T$). Due to the data collection process, the exact G_T is not available and only the approximate value $\tilde{G}$ is available. This might cause $\tilde{G}$ to be outside the set $H\Theta$, which is then known to be an ill-posed problem [Tikhonov and Arsenin 1977]. In this case, the regularization method is utilized to compute an estimate of f that is stable under small variations in the initial data.

The condition for the stability of the regularization method can be expressed mathematically as follows. $\mathcal{R}(G, \lambda)$ is defined as the regularizing operator, with λ being the regularization parameter, for the problem $Hf = G$, when only an approximation G_δ is available such that for every $\varepsilon > 0$, both $G_\delta \in \mathcal{G}$ and

$$\|G_T - G_\delta\| \leq \delta(\varepsilon), \tag{5.5}$$

imply that $max\|f_T - f_\lambda\| \leq \varepsilon$ where $f_\lambda = \mathcal{R}(G_\delta, \lambda(\delta))$. This condition requires the solution of the regularized problem to depend continuously upon the measured data and that it converges to f_T as the error $\|G_T - G_\delta\| \to 0$. In other words a slight perturbation in the measured data will not lead to large perturbation in the solution (i.e. the solution of the problem is stable).

Therefore the regularized solution f_λ, computed using $\mathcal{R}(G_\delta, \lambda)$, can be used as the solution to $Hf = G$ when only the approximation to G, G_δ, is known. According to Tikhonov [Tikhonov and Arsenin 1977], by utilizing regularization the problem is reduced to finding a suitable regularizing operator and determining the regularization parameter from supplementary information relating to the problem.

5.3 TIKHONOV-MILLER REGULARIZATION

In many applications, images are recorded to provide useful information regarding a phenomenon of interest. However, in most cases the recorded image is degraded due to either imperfections of the imaging system or the physical limitations imposed by the application or both. The resultant image is further corrupted by random noise. The imaging process, for the superresolution problem, can be expressed as

$$G = Hf + N_o, \tag{5.6}$$

where G is the Fourier transform of the degraded noisy image, H is the system transfer function of the image formation system, f is the imaged object, and N_o is the random noise. Both Tikhonov [Tikhonov and Arsenin 1977] and Hunt [Hunt 1973] proposed that a solution can be found by either minimizing $\|Hf - G\|^2$ subject to $\|Cf\|^2 \leq E^2$ where C is a linear operator (which can either be for example the identity or differentiation operator) and E^2 is the power of the object (i.e. $E^2 = \|Cf\|^2$) or minimizing $\|Cf\|^2$ subject to $\|Hf - G\|^2 \leq \epsilon^2$ where ϵ^2 is the norm of the noise. Miller [Miller 1970, Baker and Miller 1982] suggested that a solution can be found by satisfying both the following constraints: $\|Hf - G\|^2 \leq \epsilon^2$ and $\|Cf\|^2 \leq E^2$. This can be achieved by quadratically combining the above two constraints into the following constraint

$$M(f) = \|Hf - G\|^2 + \lambda\|Cf\|^2 \leq 2\epsilon^2, \tag{5.7}$$

where λ is known as the regularization parameter and is set equal to $(\frac{\epsilon}{E})^2$ [Katsaggelos 1989]. There are different methods reported for finding the regularization parameter, for example good estimates of the solution have been reported when an approximation $E^2 = \|CG\|^2$ is used [Galatsanos and Katsaggelos 1992]. Another technique to calculating regularization parameter if only ϵ^2 or E^2 is known is by finding the restored image using an initial estimate of λ and then adjusting it at each iteration based on the restored image. The process is repeated until either $\|Hf - G\|^2 \leq \epsilon^2$ or $\|Cf\|^2 \leq E^2$ is satisfied depending on which is known, ϵ^2 or E^2. Katsaggelos [Kang and Katsaggelos 1995] also proposed an iterative adaptive method which calculates the regularization parameter based on the restored image at each iteration and attempts to restore the image without any knowledge of ϵ^2 or E^2. For simplicity and since our main interest

is regarding the effects of the various constraints under study, visual inspection is used for the choice of the regularization parameter.

As mentioned previously regularization is used to transform an ill-posed problem to a well-posed one, whose solution is an approximation to that of the ill-posed problem. With regards to the superresolution problem, the Tikhonov-Miller regularization method aims at minimizing the following quadratic function

$$M(f) = \|Hf - G\|^2 + \lambda \mathcal{S}(f), \tag{5.8}$$

where λ is the regularization parameter, $\mathcal{S}(f)$ is the stabilizing operator (or regularizing function) and H is the transfer function of the imaging system, such that $H = \xi \mathcal{F} \gamma$, where γ is the space domain constraint, ξ is the frequency domain constraint and $\mathcal{F}$ is the forward Fourier transform. Minimizing the first term of the above equation, $\|Hf - G\|^2$, aims at preserving the properties of the original object and maintaining consistency with the available data G. The latter part of the equation, $\mathcal{S}(f)$, enforces smoothness of the solution. Therefore, the regularization parameter, λ, determines the trade off between the fidelity of the solution to the data G and the numerical stability. A large λ provides a stable solution with poor fidelity, whereas a small λ provides insufficient regularization to overcome the ill-posedness [Katsaggelos *et al.* 1991]. The stabilizing operator, $\mathcal{S}$, incorporates prior knowledge of the smoothness of f [Verveer and Jovin 1997]. When there is no prior knowledge of f_T available, $\mathcal{S}(f)$ is set to $\|f\|^2$ and thus the norm of the solution is minimized (representing a bound on the energy of the image); this is known as the Tikhonov-Miller regularization method.

The regularization method can be incorporated within the steepest descent method as follows. The steepest descent iteration (with steplength equal to $\frac{1}{2}$) is

$$f^{k+1} = f^k + \frac{1}{2} r^k \tag{5.9}$$

$$\begin{aligned} \text{where} \quad r^k &= -\nabla_f M(f^k), \\ &= -2H^* H f^k + 2H^* G - \lambda \nabla_f \mathcal{S}(f^k) \end{aligned} \tag{5.10}$$

where $H^* = \gamma \mathcal{F}^{-1} \xi$. Substituting Eqn. 5.10 into Eqn. 5.9 we have

$$f^{k+1} = f^k + H^* G - H^* H f^k - \frac{1}{2} \lambda \nabla_f \mathcal{S}(f^k). \tag{5.11}$$

Using Tikhonov-Miller regularization, i.e. $\mathcal{S}(f) = \|f\|^2$, this leads to

$$f^{k+1} = f^k + H^* G - H^* H f^k - \lambda f^k. \tag{5.12}$$

The Tikhonov-Miller regularization was applied to the bandlimited head model,

shown in Fig. 3.4(b). Fig. 5.1 shows the ISNR curves obtained from applying regularization for different values of the regularization parameter. The effect of the regularization is very obvious from the levelling off of the ISNR curves, making them less sensitive to the number of iterations performed. In other words, the solution is more stable. Note that $\lambda = 0$ corresponds to the G-P solution (no regularization), and that with $\lambda = 0.01$, the maximum ISNR achieved is only slightly less that for the other cases.

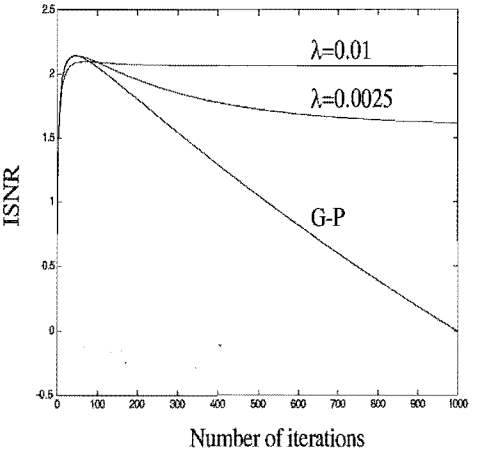


Figure 5.1 ISNR curves using Tikhonov-Miller regularization with $\lambda = 0, 0.0025, 0.01$.

5.4 MODEL-BASED REGULARIZATION

This section presents a modified version of the Tikhonov-Miller regularization, where a model of the solution is incorporated within the algorithm. Improvements in the results obtained depend on the precision of the model being used. The idea of using a prior model is to introduce properties of the original image in the superresolving process. The more precise the model is (compared to the original image), the greater the improvement in the solution. This method aims at minimizing the following quadratic function [Barakat *et al.* 1997]

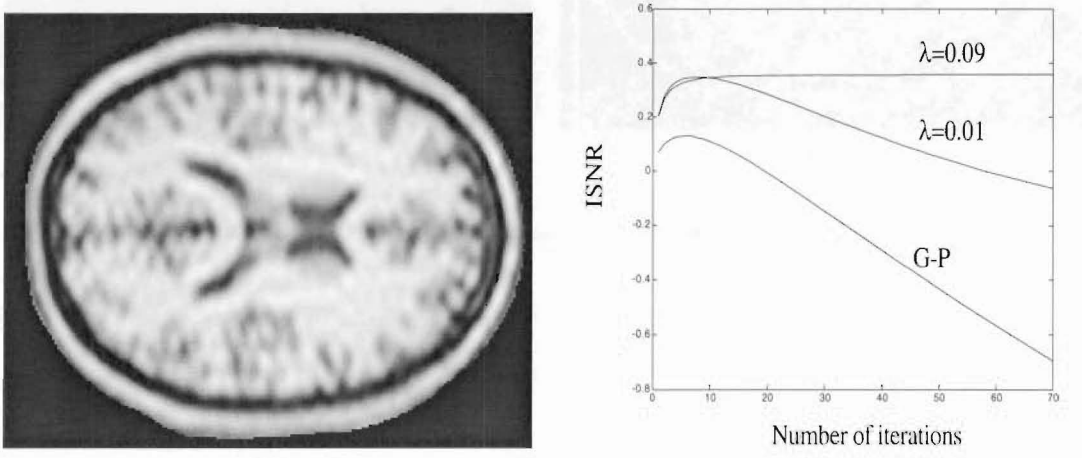
$$M(f) = \|Hf - G\|^2 + \lambda\|f - f_{model}\|^2. \tag{5.13}$$

Following the procedure in the previous section, the steepest descent method may be written as

$$f^{k+1} = f^k + H^*G - H^*Hf^k - \lambda(f^k - f_{model}). \tag{5.14}$$

It is obvious that if the model used lacks precision, the errors in the solution may exceed those occurring without utilizing the model.

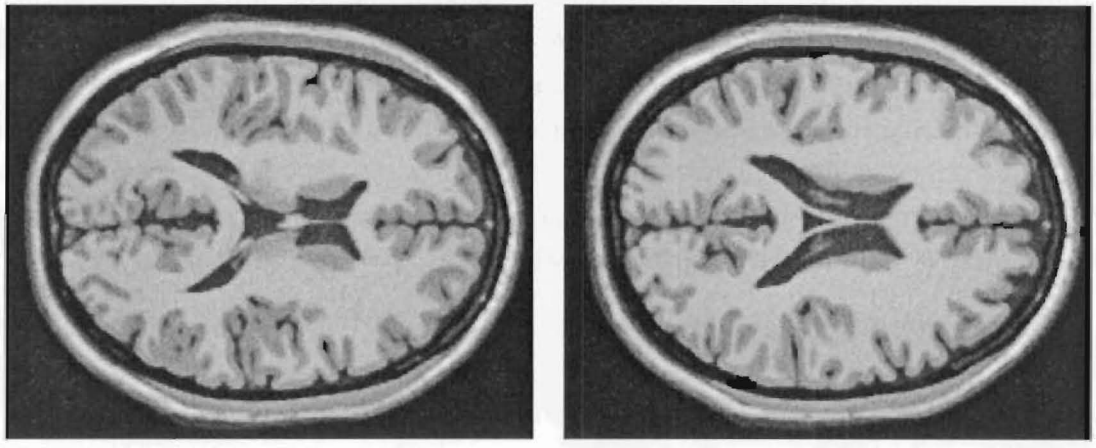
The model-based regularization method was applied on the bandlimited brain image, shown in Fig. 3.5(b). Fig. 5.2(b) shows the ISNR curves obtained from applying the model-based regularization for different values of the regularization parameter. The bandlimited noisy brain image was used as the model. The aim of regularizing the superresolution problem for the brain image, i.e. levelling off of the ISNR curves, was achieved. Fig. 5.2(a) shows the recovered image (at iteration 70) obtained for $\lambda = 0.09$. When comparing the high resolution image, shown in Fig. 3.5(a), the bandlimited image and the recovered image, no visual improvement was observed. The reason for this is that the model used provided no additional information to the problem since the initial data G and the model used were both of the low resolution image.



(a) (b)

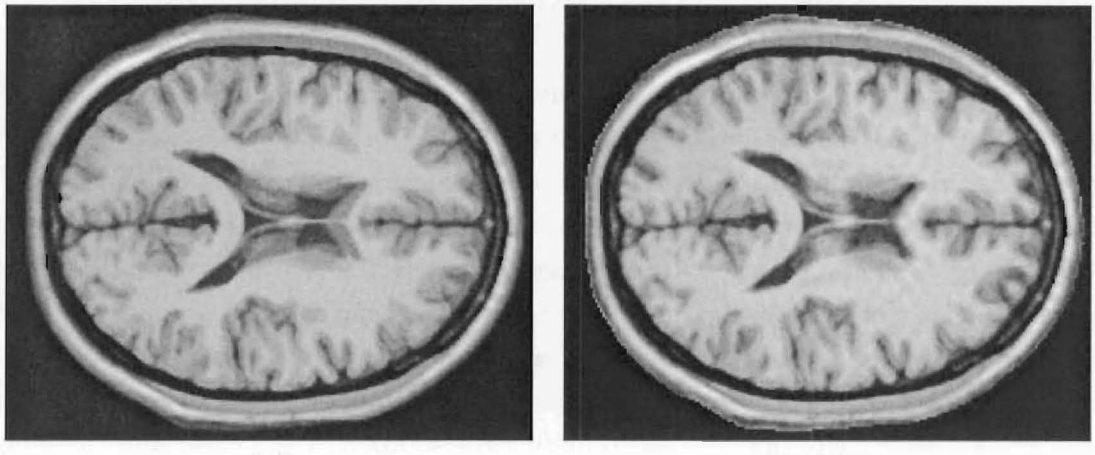
Figure 5.2 Model-based regularization: (a) recovered brain image with $\lambda = 0.09$; (b) ISNR curves for $\lambda = 0, 0.01, 0.09$.

A more practical situation utilizing this method arises if a low resolution image of a slice is acquired and the images of the two adjacent slices (above f_1, and below f_2) were of a higher resolution. This case requires less imaging time than that of acquiring high resolution images of all the slices. The model utilized could then be the average of the two high resolution images. It is important to note that the high resolution slices themselves are not used as the model to avoid the incorporation of unwanted details (present in the high resolution slices only) in the solution. If the distance between the slices is small, then the slices will have some dependency between them, and f_1 and f_2 will provide some extra information to finding the solution. Fig. 5.3(a) and (b) show the two high resolution slices, f_1 and f_2 respectively and Fig. 5.4(a) shows the average of the two slices. Fig. 5.4(b) shows the recovered brain image (at iteration 100) with the average image being the model. In this case the availability of a better model has lead to some visual improvement. Careful choice of λ is important here, since a large value of λ will result in the solution being close to the model utilized and a very small value of λ will hardly provide any regularization.



(a) (b)

Figure 5.3 Model-based regularization: (a) slice 1; (b) slice 2.



(a) (b)

Figure 5.4 Model-based regularization: (a) average slice; (b) recovered image with $\lambda = .04$.

5.5 SMOOTHNESS CONSTRAINT

In many practical applications, the object being imaged has regions of equal intensity (e.g. in MRI images, pixels belonging to the same tissue), however due to many factors discussed previously, the acquired image is degraded and suffers from Gibb's oscillations. A smoothness constraint forces neighbouring pixels in the solution to have similar intensity values. Since this constraint requires some neighbourhood process, the most common operator utilized is the differential operator, where for example if the difference between two adjacent pixels is large, the application of the smoothness constraint aims at reducing this difference in the solution [Sementilli *et al.* 1993]. However, a problem arises at edges where the difference between the adjacent pixels is large, smoothing out of these edges reduces high spatial frequency content, which conflicts with the aim of superresolution [Kang and Katsaggelos 1995]. This problem is addressed in section 5.6, where edge-preserving regularization is discussed.

A different choice for $\mathcal{S}(f)$ in Eqn. 5.8 leads to a smoothness-based regularization. By basing $\mathcal{S}(f)$ on the differential operator, thus $\mathcal{S}(f) = \|\nabla f\|^2$, smoothness of the solution is maximized. The steepest descent method is then utilized to minimize the following function

$$M(f) = \|Hf - G\|^2 + \lambda\|\nabla f\|^2. \tag{5.15}$$

By comparison with Eqns. 5.9-5.11, the Tikhonov-Miller regularization with smoothness constraint can be written as

$$f^{k+1} = f^k + H^*G - H^*Hf^k - \frac{1}{2}\lambda\nabla_f\mathcal{S}(f). \tag{5.16}$$

As nonlinear forms of the smoothness constraint are to be considered in section 5.6, we assume a more general form for the regularization function

$$\mathcal{S}(f) = \int_{\mathcal{V}} \Phi(\nabla f \cdot \nabla f)dv = \int_{\mathcal{V}} \Phi(|\nabla f|^2)dv, \tag{5.17}$$

where $\mathcal{V}$ is a closed region of space bound by a regular boundary $\mathcal{B}$, and let $\Psi(\eta) = \Phi(\eta^2)$ which is known as the penalty function, where $\eta = |\nabla f|$. As mentioned previously the regularization function, $\mathcal{S}(f)$, penalizes the irregularities according to the *a priori* smoothness constraint encoded in it, where the encoding is achieved by utilizing a penalty function which is a penalty against irregularities in f.

To compute the gradient of $\mathcal{S}(f)$, a perturbation δf in the function f is considered, with the linearization of the regularization function given by (utilizing Taylor's theorem

and neglecting second order terms)

$$\begin{aligned} L(f)\delta f &= \mathcal{S}(f+\delta f) - \mathcal{S}(f) \\ &= \int_{\mathcal{V}} \Phi(\nabla(f+\delta f)\cdot\nabla(f+\delta f))dv - \int_{\mathcal{V}} \Phi(\nabla f\cdot\nabla f)dv \\ &= \int_{\mathcal{V}} [\Phi(\nabla f\cdot\nabla f + 2\nabla\delta f\cdot\nabla f) - \Phi(\nabla f\cdot\nabla f)]dv \\ &= 2\int_{\mathcal{V}} \Phi'(\nabla f\cdot\nabla f)\nabla\delta f\cdot\nabla f dv. \end{aligned} \tag{5.18}$$

Applying the identity $\nabla(ab) = a\nabla b + b\nabla a$, and Green's first theorem (with n being the outward normal to the boundary $\mathcal{B}$) [Stratton 1941], we have

$$\begin{aligned} L(f)\delta f &= 2\{\int_{\mathcal{V}} \nabla(\delta f\Phi'(\nabla f\cdot\nabla f))\cdot\nabla f dv - \int_{\mathcal{V}} \delta f\nabla\Phi'(\nabla f\cdot\nabla f)\cdot\nabla f dv\} \\ &= 2\{\int_{\mathcal{B}} \delta f\Phi'(\nabla f\cdot\nabla f)\frac{\delta f}{\delta n}ds - \int_{\mathcal{V}} \delta f\Phi'(\nabla f\cdot\nabla f)\nabla^2 f dv - \int_{\mathcal{V}} \delta f\nabla\Phi'(\nabla f\cdot\nabla f)\cdot\nabla f dv\}. \end{aligned}$$

Applying expansion of the divergence ($\nabla\cdot(c\nabla d) = \nabla c\cdot\nabla d + c\nabla^2 d$), and then setting $\frac{\delta f}{\delta n} = 0$, for $(x,y)\in S$ (Neumann boundary condition), we have

$$\begin{aligned} L(f)\delta f &= 2\{\int_{\mathcal{B}} \delta f\Phi'(\nabla f\cdot\nabla f)\frac{\delta f}{\delta n}ds - [\int_{\mathcal{V}} \delta f\nabla\cdot(\Phi'(\nabla f\cdot\nabla f)\nabla f)dv \\ &\quad - \int_{\mathcal{V}} \delta f\nabla\Phi'(\nabla f\cdot\nabla f\cdot\nabla f)dv] - \int_{\mathcal{V}} \delta f\nabla\Phi'(\nabla f\cdot\nabla f)\cdot\nabla f dv\} \\ &= -2\int_{\mathcal{V}} \delta f\nabla\cdot(\Phi'(\nabla f\cdot\nabla f)\nabla f)dv. \end{aligned} \tag{5.19}$$

Since the linearization of $\mathcal{S}(f)$ is defined as the inner product of the gradient and the perturbation (i.e. $L(f)\delta f = < \mathcal{S}'(f), \delta f >$), the gradient is

$$\mathcal{S}'(f) = -2\nabla\cdot(\Phi'(|\nabla f|^2)\nabla f). \tag{5.20}$$

where the notation has changed from $\nabla_f\mathcal{S}$ to $\mathcal{S}'(f)$ to avoid confusion with the spatial derivatives.

For the Tikhonov-Miller method with smoothness constraint, the stabilizing operator $\mathcal{S} = \|\nabla f\|^2$, therefore $\Psi(\eta) = \Phi(\eta^2) = \eta^2$, where $\eta = |\nabla f|$ and so $\Phi(|\nabla f|^2) = |\nabla f|^2$, and the gradient for the smoothness constraint is

$$\mathcal{S}'(f) = -2\nabla\cdot(\nabla f) = -2\nabla^2 f, \tag{5.21}$$

Since for the smoothness constraint $\Psi(\eta) = \eta^2$, the smoothness constraint is also known as the standard quadratic regularization. At this point, it is important to define the interaction function $\mathcal{I}(\eta) = \frac{\Psi'(\eta)}{2\eta}$ which determines the interaction between neighbouring pixels, and $|\Psi'(\eta)| = |2\eta\mathcal{I}(\eta)|$ is the strength at which the regularization function

performs smoothing i.e the amount of influence imposed as the gradient increases. In order for edges to be preserved, the following condition must be satisfied [Li 1995]:

$$\lim_{\eta\to\infty} |\Psi'(\eta)| = \lim_{\eta\to\infty} |2\eta\mathcal{I}(\eta)| = C, \tag{5.22}$$

where $C \geq 0$. Therefore if $C = 0$, smoothing at discontinuities is absolutely not allowed, however if $C > 0$ limited smoothing is allowed at discontinuities. To satisfy Eqn. 5.22, $\mathcal{I}(\eta) \to 0$ as $\eta \to \infty$.

The functions $\Psi(\eta)$, $\Psi'(\eta)$ and the corresponding interaction function $\mathcal{I}(\eta)$ are shown in Fig. 5.5 (a),(b) and (c) respectively. In this case the smoothing strength $|\Psi'(\eta)| = |2\eta|$ increases as η increases, and $|2\eta| \to \infty$ as $\eta \to \infty$ therefore boundless smoothing is allowed, and edges are oversmoothed. Note that the interaction function is unity everywhere, therefore Eqn. 5.22 is not satisfied resulting in oversmoothing.

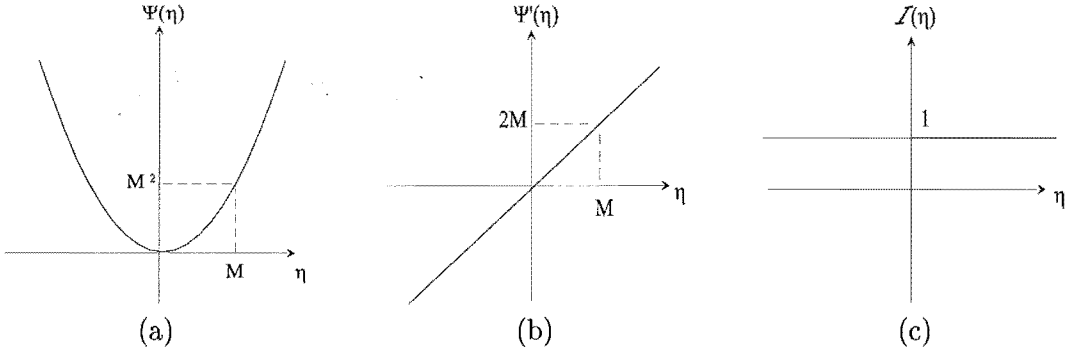


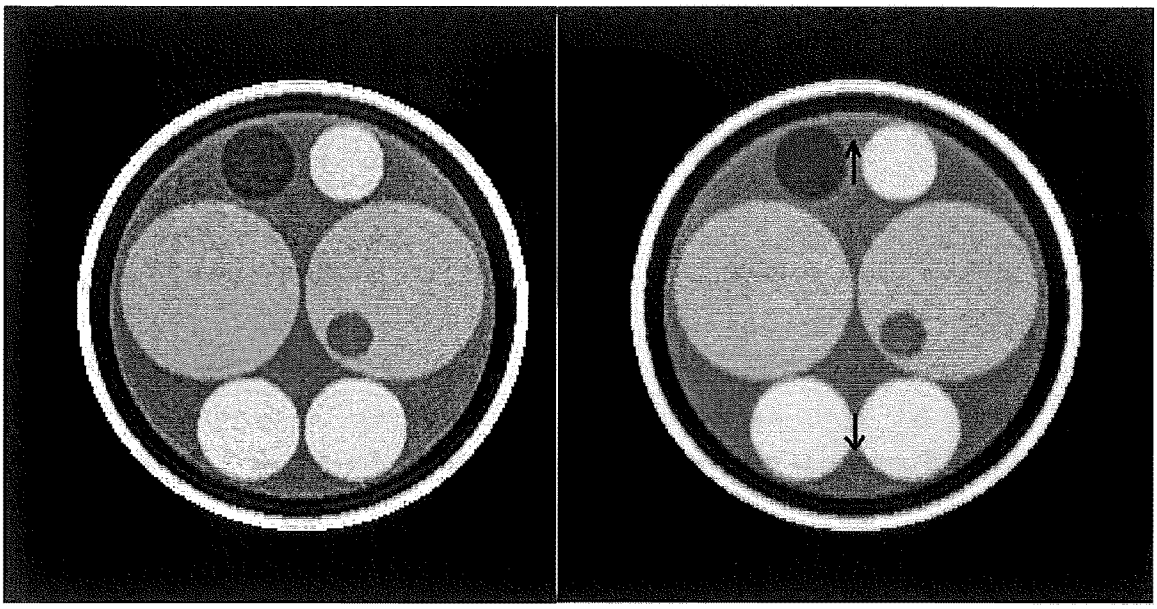
Figure 5.5 Smoothness constraint: (a) $\Psi(\eta)$; (b) $\Psi'(\eta)$; (c) $\mathcal{I}(\eta)$.

The smoothness constraint was applied on the bandlimited head image. Fig. 5.6(a) and (b) show the recovered head images (at iteration 1000) utilizing the smoothness operator with $\lambda = 0.01$ and $\lambda = 0.5$ respectively. Note that for small values of λ, the problem is regularized but with very little enforcement of the smoothness constraint. However when λ is large, (i.e. $\lambda = 0.5$), the effect of the constraint is very obvious from the smoothed out edges of the recovered image, however the image starts 'breaking up' (regions indicated by the arrows). The reason for the image breaking up is the nonconvergence of the algorithm i.e. the steplength, α^k discussed in section 3.5, must be controlled carefully (to achieve convergence) in order to avoid an unexpected increase in f^k when α^k is too large for a large value of λ [Himmelblau 1972]. For an arbitrary steplength value Eqn. 5.16 becomes

$$f^{k+1} = f^k + 2\alpha^k H^*G - 2\alpha^k H^*Hf^k - \alpha^k\lambda\nabla_f\mathcal{S}(f), \tag{5.23}$$

and for the smoothness constraint we have

$$f^{k+1} = f^k + 2\alpha^k H^*G - 2\alpha^k H^*Hf^k + 2\alpha^k\lambda\nabla^2 f. \tag{5.24}$$



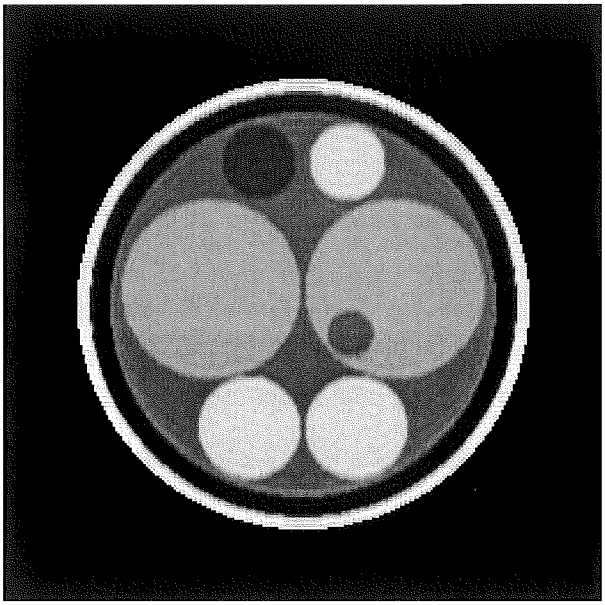
(a) (b)

Figure 5.6 Recovered image obtained from applying the smoothness constraint: (a) with $\lambda = 0.01, \alpha^k = 0.5$; (b) recovered head image with $\lambda = 0.5, \alpha^k = 0.5$.

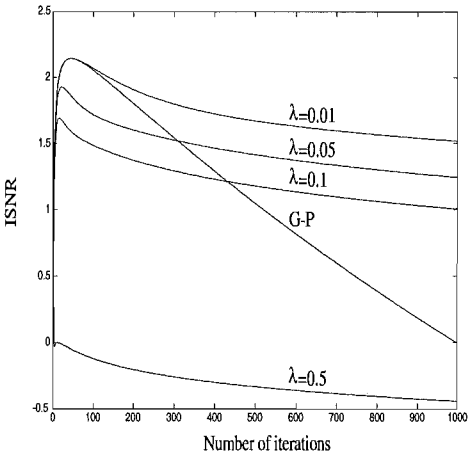
To avoid this problem the steplength, α^k, was set to 0.05 instead of 0.5 for $\lambda = 0.01$ (a decrease by a factor of 10) and $\alpha^k = 0.005$ for $\lambda = 0.5$ (a decrease by a factor of 100). The recovered image (at iteration 1000) for $\lambda = 0.5$ and $\alpha^k = 0.005$ is shown in Fig. 5.7(a) which shows the edges smoothed out, Gibb's oscillation reduced and the problem of the breaking up of the image is avoided. In order to compare the ISNR curves obtained from applying the smoothness constraint to that of the G-P algorithm, the ISNR curves for the smoothness constraint for $\lambda = 0.01, 0.05, 0.1, 0.5$ were plotted in steps of 10,10,100,100 respectively as shown in Fig. 5.7(b). The low ISNR curve for the case when $\lambda = 0.5$, is due to the loss of the sharp edges even though Gibb's oscillations have been reduced. It is important to note that once the steplength is reduced sufficiently such that convergence is achieved, further reduction would also result in convergence and at effectively the same rate (i.e. same ISNR curve). Faster convergence would be obtained with the application of the conjugate gradient method here (discussed in section 3.6).

5.6 EDGE PRESERVING METHODS

In image processing, regularization is used to smooth out the solution, to remove noise or to enhance features of the image. While it is important to recover the image over smooth patches, the transitions between these patches are also important. One of the main disadvantages of utilizing the Tikhonov-Miller regularization is that it penalizes discontinuities in the image and is unable to recover sharp edges. It tends to



(a)



(b)

Figure 5.7 Smoothness constraint: (a) recovered image with $\lambda = 0.5, \alpha^k = 0.005$; (b) ISNR curve.

attenuate high frequencies, in opposition to the aims of superresolution. The most important property desired by the regularizing algorithms is the edge preserving ability. The nonlinear function, Ψ may be used to protect these discontinuities from being suppressed by the smoothness constraint by choosing Ψ' such that small differences between neighbouring pixels over a region are suppressed and highly localized differences of large magnitudes may be retained. There are many different proposed forms for the nonlinear operator. The following subsections 5.6.1 to 5.6.3 discuss some of them.

5.6.1 Total variation regularization

Up to this point, the regularization methods studied in general have not preserved edges in the presence of noise in the data. An edge-preserving regularization method was proposed by [Rudin *et al.* 1992] which minimizes the total variation (TV) of the image [Connolly *et al.* 1999]

$$\|f\|_{TV} = \int |\nabla f| dxdy = \int (\nabla f \cdot \nabla f)^{\frac{1}{2}} dxdy, \tag{5.25}$$

while preserving some fit to the original measured data. The total variation method preserves the location of the edges without penalizing smooth functions. It searches for an approximate solution matching the measured data that has minimal total variation, but with no particular bias towards either a discontinuous or smooth solution. Applying the total variation method to the superresolution problem, with $\mathcal{S}(f) = \|f\|_{TV}$, Eqn.

5.8 can be written as

$$M(f) = \|Hf - G\|^2 + \lambda \|f\|_{TV}. \tag{5.26}$$

The total variation method seeks an approximation to the observed data with the least total variation, with no any preference for either continuous or discontinuous functions. To illustrate this Fig. 5.8 shows a set of functions which are required to satisfy the conditions $f(0) = a$ and $f(1) = b$. All functions f_1, f_2, f_3 and f_4 satisfy the above conditions and would be acceptable to $\min_f TV(f)$ since $TV(f_1) = TV(f_2) = TV(f_3) = TV(f_4) = b - a$ [Strong and Chan 1996b, Chan and Mulet 1996].

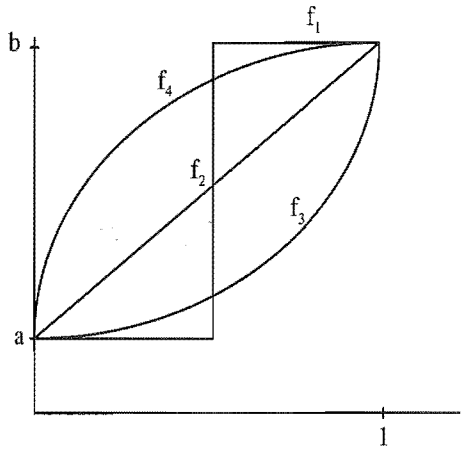


Figure 5.8 All functions f_1, f_2, f_3 and f_4 have a total variation of $b - a$.

The penalty function for the total variation regularization is $\Psi(\eta) = \Phi(\eta^2) = |\eta|$, as shown in Fig. 5.9(a), therefore $\Psi'(\eta) = \frac{\eta}{|\eta|}$, shown in Fig. 5.9(b), and the edge preserving condition on the smoothing function $|\Psi'|$, $\lim_{\eta \to \infty} |\Psi'(\eta)| = 1$ (or in other words $\mathcal{I}(\eta) \to 0$ as $\eta \to \infty$) is satisfied, with $C = 1$ indicating a bounded smoothing at discontinuities.

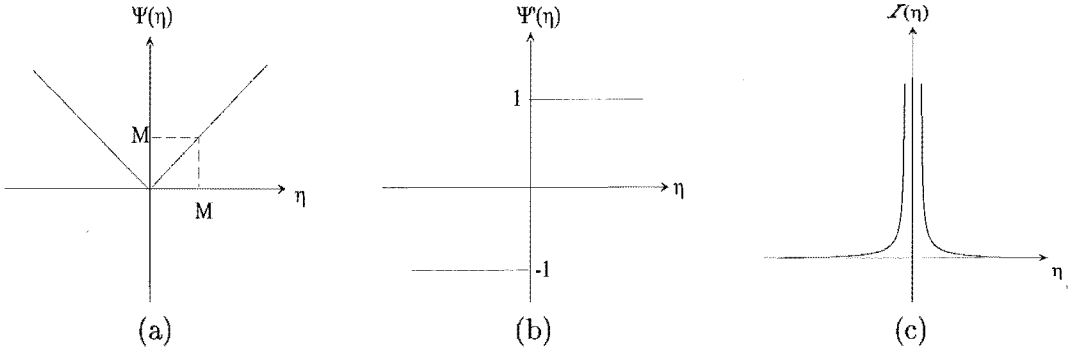


Figure 5.9 Total variation regularization: (a) $\Psi(\eta)$; (b) $\Psi'(\eta)$; (c) $\mathcal{I}(\eta)$.

Setting $\Psi(\eta) = \Phi(\eta^2) = |\eta|$ so that $\Phi(|\nabla f|^2) = |\nabla f|$, and referring back to Eqn.

5.20, the gradient for the total variation method is

$$\mathcal{S}'(f) = -\nabla \cdot (\frac{\nabla f}{|\nabla f|}). \tag{5.27}$$

For some points it is possible that $|\nabla f| = 0$ and thus $\mathcal{S}'(f)$ may be undefined. A common technique used to overcome this difficulty is to replace $|\nabla f|$ by $\sqrt{|\nabla f|^2 + \beta}$, where β is a small positive number [Vogel and Oman 1996].

The total variation method was applied on the bandlimited head image. Fig. 5.10(a) shows the recovered head image (at iteration 1000) using the total variation method with $\lambda = 0.05$ and $\alpha^k = 0.05$, which shows a significant edge enhancement compared to the bandlimited image (and to the recovered image using the smoothness constraint, Fig. 5.7(a)). In addition to the edge enhancement, the total variation method clears the Gibb's oscillation and reduces noise. The clearance of Gibb's oscillation is due to the fact that all oscillations have a high total variation, therefore by applying the total variation method, which aims at reducing the total variation, oscillations are smoothed out. However if a very high level of total variation is applied (i.e. λ is very large), the removal of oscillations and noise is accompanied with a loss of detail. The amount of total variation (i.e. λ) should be carefully chosen such that it is high enough to remove oscillations but low enough to preserve details [Strong and Chan 1996a, Strong 1997]. Fig. 5.10(b) shows the corresponding ISNR curve compared to the ISNR curve obtained from applying the G-P algorithm. The ISNR curve obtained from the total variation method is considerably levelled off (and of a much higher level) compared to the ISNR curve obtained from the G-P algorithm, which indicates an improvement in stability.

Since the whole idea of using the total variation method is to recover edges, another way of studying the performance of this method is to compare the gradients of both the bandlimited head image and the recovered image using the total variation method as shown in Figs. 5.11 (a) and (b) respectively, where the images are scaled to improve display. It is clear from these images that the edges are much sharper using the total variation method.

The total variation method is useful when applied on 'blocky' images since they have a small total variation [Dobson and Santosa 1996], however when applied on highly oscillatory images (i.e. that have a large total variation), important details are removed. To illustrate this, the total variation method was also applied on the bandlimited brain image with $\lambda = 0.5$, resulting in the recovered image shown in Fig. 5.12. Note the the slight visual enhancement of the central part of the image (with has a relatively low total variation) but the undesirable blocky effect and the loss of details of features surrounding the central part (which have a high total variation). Better results would be expected on a less variable object such as a leg. Also later in section 7.5, it is shown that the total variation method is useful when applied to a 'region of interest' of the

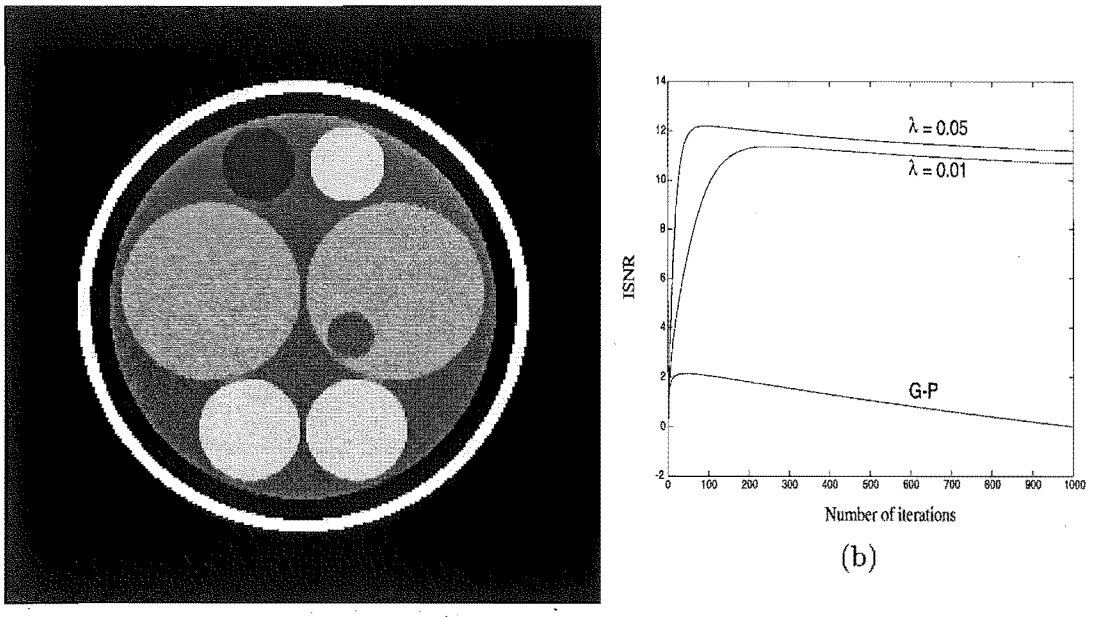


Figure 5.10 Total variation method (a) recovered head image with $\lambda = 0.05$, $\alpha^k = 0.05$; (b) ISNR curve.

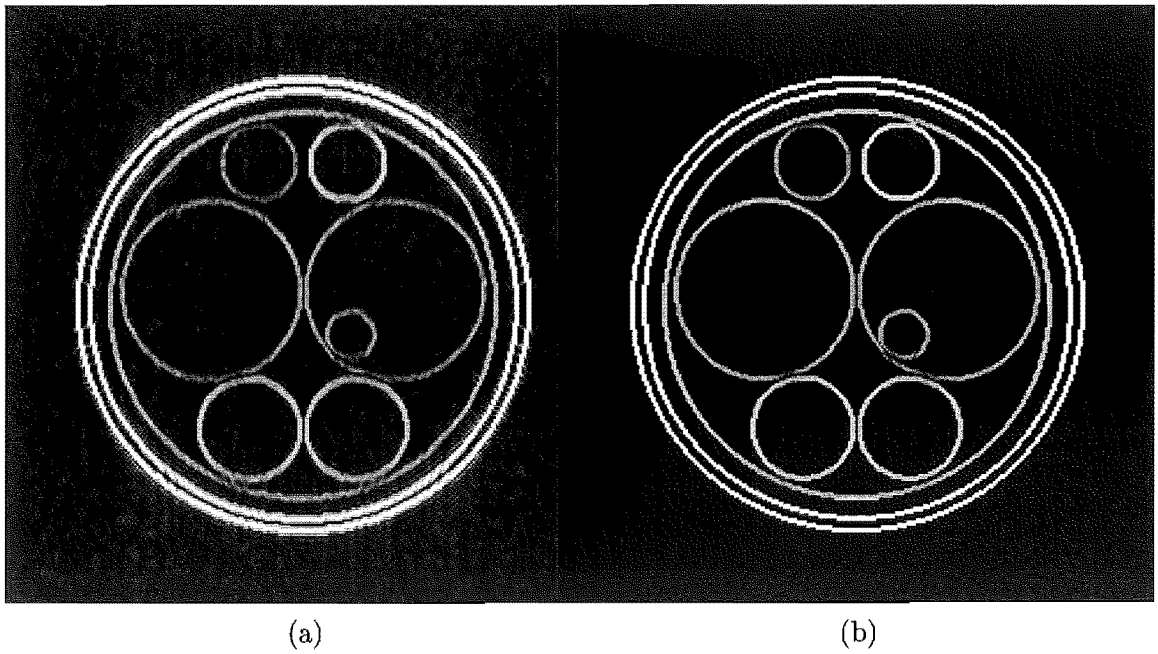


Figure 5.11 Images of $|\nabla f|$ where: (a) f is the bandlimited head image; (b) f is the recovered image obtained from utilizing the total variation method with $\lambda = 0.05$, $\alpha^k = 0.05$.

brain image (i.e. when application of the constraint is restricted to a particular feature such as a tumour).

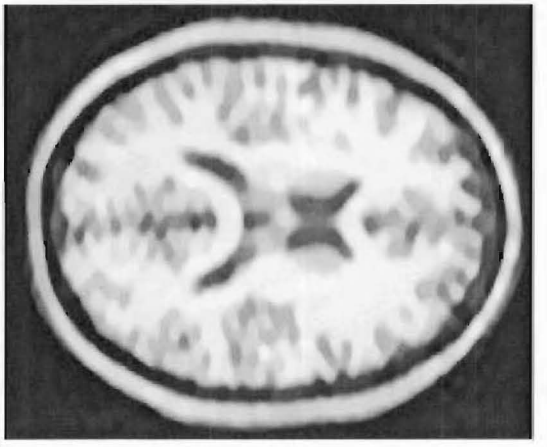


Figure 5.12 Recovered brain image from applying total variation method with $\lambda = 0.5$.

5.6.2 Edge preserving regularization using the Huber function

A very well known penalty function is the Huber penalty function. It was applied by Stevenson and Delp to the problems of surface [Stevenson *et al.* 1991] and curve [Stevenson and Delp 1990] reconstruction and by Shulman and Hervé [Shulman and Hervé 1989] in regularizing discontinuous flow fields, but was originally motivated by the work done by Huber in robust statistics [Huber 1981].

The Huber penalty function is quadratic near zero but is linear beyond a threshold value M [Delaney and Bresler 1998]. This function acts as a smoothness constraint for small values of the gradient and as a total variation regularization for large gradient values. Therefore, this penalty allows for images that are a composite of both sharp edges and smoothly varying regions. Since it is a convex function, the penalty results in a convex functional to be minimized giving a unique global minimizer to the problem.

The Huber penalty function can be mathematically defined as follows

$$\Psi(\eta) = \begin{cases} \eta^2, & |\eta| \leq M \\ M^2 + 2M(|\eta| - M), & |\eta| > M. \end{cases} \tag{5.28}$$

and is illustrated in Fig. 5.13(a). The first derivative of the Huber penalty function is 2η for $|\eta| \leq M$ and $\frac{2M\eta}{|\eta|}$ elsewhere, therefore the interaction function is 1 for $|\eta| \leq M$ and $\frac{M}{|\eta|}$ elsewhere. Note that for $|\eta| \leq M$, $\lim_{\eta \to \infty} |\Psi'(\eta)|$ is a constant and as such allows bounded but non-zero smoothing at discontinuities [Li 1998]. Figs. 5.13 (b) and (c) show the smoothing strength and interaction function respectively. Note that for $\eta \leq M$, $\mathcal{I}(\eta)$ is the same as that of the smoothness constraint (shown in Fig. 5.5(c))

and for $\eta > M$, $\mathcal{I}(\eta)$ is the same as the that of the total variation method (shown in Fig. 5.9(c)).

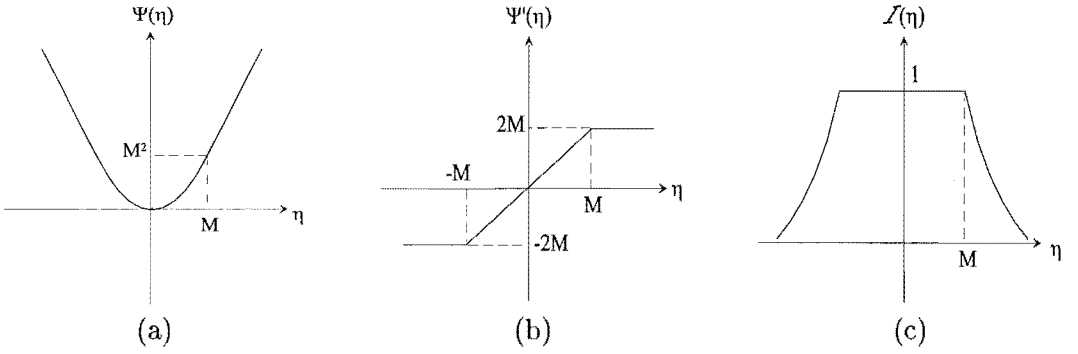


Figure 5.13 Huber function: (a) $\Psi(\eta)$; (b) $\Psi'(\eta)$; (c) $\mathcal{I}(\eta)$.

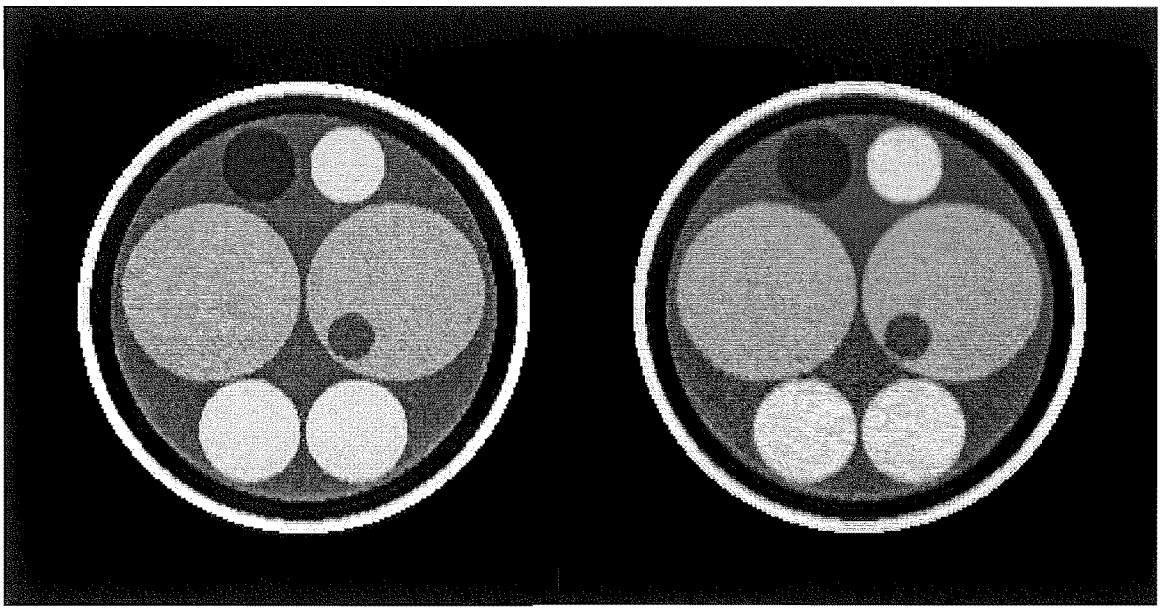
The gradient for the regularization method utilizing the Huber function, using Eqn. 5.20, can then be formulated as

$$\mathcal{S}'(f) = \begin{cases} -2\nabla^2 f, & |\nabla f| \leq M \\ -2M\nabla \cdot (\frac{\nabla f}{|\nabla f|}), & |\nabla f| > M. \end{cases} \tag{5.29}$$

The Huber penalty function was applied on the bandlimited head data, with $\lambda = 0.5$ and threshold values $M = 0.001, 1$. Comparing the recovered images for the two different threshold values, we note that Fig. 5.14(a) (for $M = 0.001$) has much sharper edges than Fig. 5.14(b) since total variation is applied strongly on the image and the smoothness constraint is weakly applied. However for both cases Gibb's oscillation is significantly reduced.

5.6.3 Edge preserving regularization using the truncated quadratic function

Up to this point all the regularization methods studied apply a convex penalty function, and do not have the problems found when using nonconvex penalty functions such as the solution being dependant on the initialization as only a local minima can be guaranteed [Li 1998]. In addition nonconvex penalty functions may not necessarily regularize the problem [Connolly *et al.* 1999], since if the penalty function is nonconvex, the problem of minimizing a nonconvex functional is itself ill-posed and the solution will no longer vary continuously with a change in input [Stevenson *et al.* 1994]. However, for images consisting of smooth regions with sharp boundaries, it has been found that nonconvex penalty functions may be better in some cases for recovering edges since edges are less severely penalized than when utilizing convex penalty functions [Delaney and Bresler 1998]. Note that there are many other different penalty functions (other



(a) (b)

Figure 5.14 Huber function: Recovered image using the Huber function with $\lambda = 0.5$, $\alpha^k = 0.0005$ and : (a) $M = 0.001$; (b) $M = 1$.

than those discussed here), convex and nonconvex, found in various literature [Geman and Reynolds 1992, Lorette *et al.* 1997].

Blake and Zisserman [Blake and Zisserman 1987] proposed a nonconvex penalty function, known as the truncated quadratic regularization method. This penalty function can be mathematically defined as follows

$$\Psi(\eta) = \begin{cases} \eta^2, & |\eta| \leq M \\ M^2, & |\eta| > M, \end{cases} \tag{5.30}$$

and is illustrated in Fig. 5.15(a), where M is the threshold value. Note that the penalty function is quadratic near 0 and acts as a smoothness constraint but is flat beyond the threshold value to avoid penalizing large gradients and allow sharp edges in the solution [Bouman and Sauer 1993]. The interaction function, shown in Fig. 5.15(c), $\mathcal{I}(\eta) = 1$ if $|\eta| \leq M$ and 0 elsewhere, and the smoothing strength $|\Psi'(\eta)| = 0$ when $|\eta| > M$, shown in Fig. 5.15(b), which prohibits smoothing for gradient values greater than the threshold M. Note that for $\eta \leq M$, $\mathcal{I}(\eta)$ is the same as that of the smoothness constraint (shown in Fig. 5.5(c)).

The gradient for the truncated quadratic regularization method, using Eqn. 5.20, can then be formulated as

$$\mathcal{S}'(f) = \begin{cases} -2\nabla^2 f, & |\nabla f| \leq M \\ 0, & |\nabla f| > M. \end{cases} \tag{5.31}$$

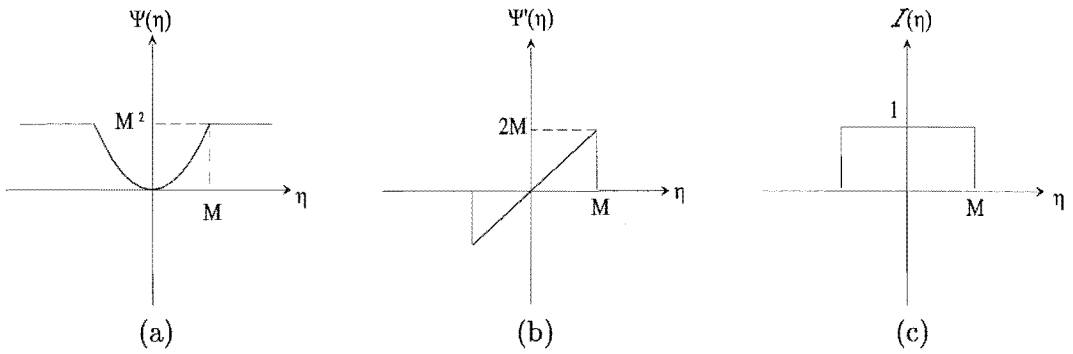
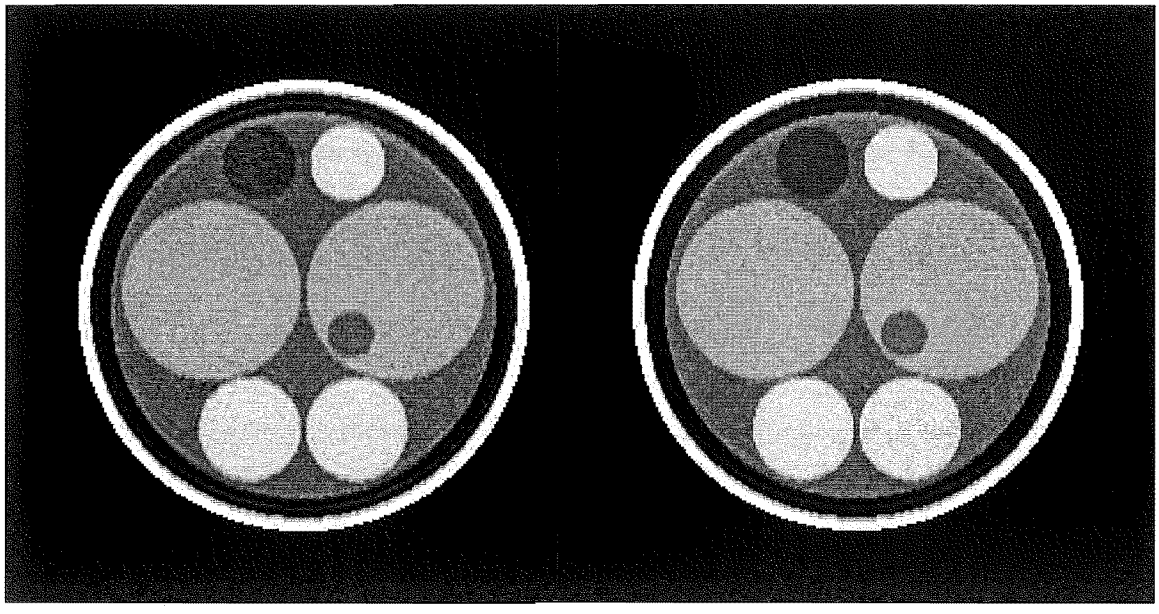


Figure 5.15 Truncated quadratic function: (a) $\Psi(\eta)$; (b) $\Psi'(\eta)$; (c) $\mathcal{I}(\eta)$.

Regularization utilizing the truncated quadratic function was applied on the bandlimited noisy head image and the recovered images for $M = 0.001$ (at iteration 45) and $M = 0.05$ (at iteration 12) are shown in Figs. 5.16(a) and (b) respectively. For the case $M = 0.05$, the recovered image is smoothed out enough to significantly reduce Gibb's oscillations while still preserving the edges of the important features as opposed to the case when $M = 0.001$ where no reduction in Gibb's oscillation is obvious (very similar to the image obtained form applying only the G-P algorithm (see Fig. 3.6(a)).



(a) (b)

Figure 5.16 Recovered head image using the truncated quadratic function with $\lambda = 0.1$: (a) $M = 0.001$; (b) $M = 0.05$.

Fig. 5.17(a) and (b) show the areas (in white) where smoothing was applied for the threshold value $M = 0.001$ and $M = 0.05$ respectively and Fig. 5.17(c) shows the corresponding ISNR curve obtained from applying the truncated quadratic function. Note the nonconvergence of the algorithm, i.e. problem is not regularized, requiring

early termination of the iterations to obtain a regularized solution. The rate of divergence after about 500 iterations seems comparable with that obtained from the G-P method. Additional constraints (e.g. energy constraints) may help to stabilize the problem. Note that a reduction in steplength is not necessary and the occurrence of a local minima does not appear to be a difficulty in this example.

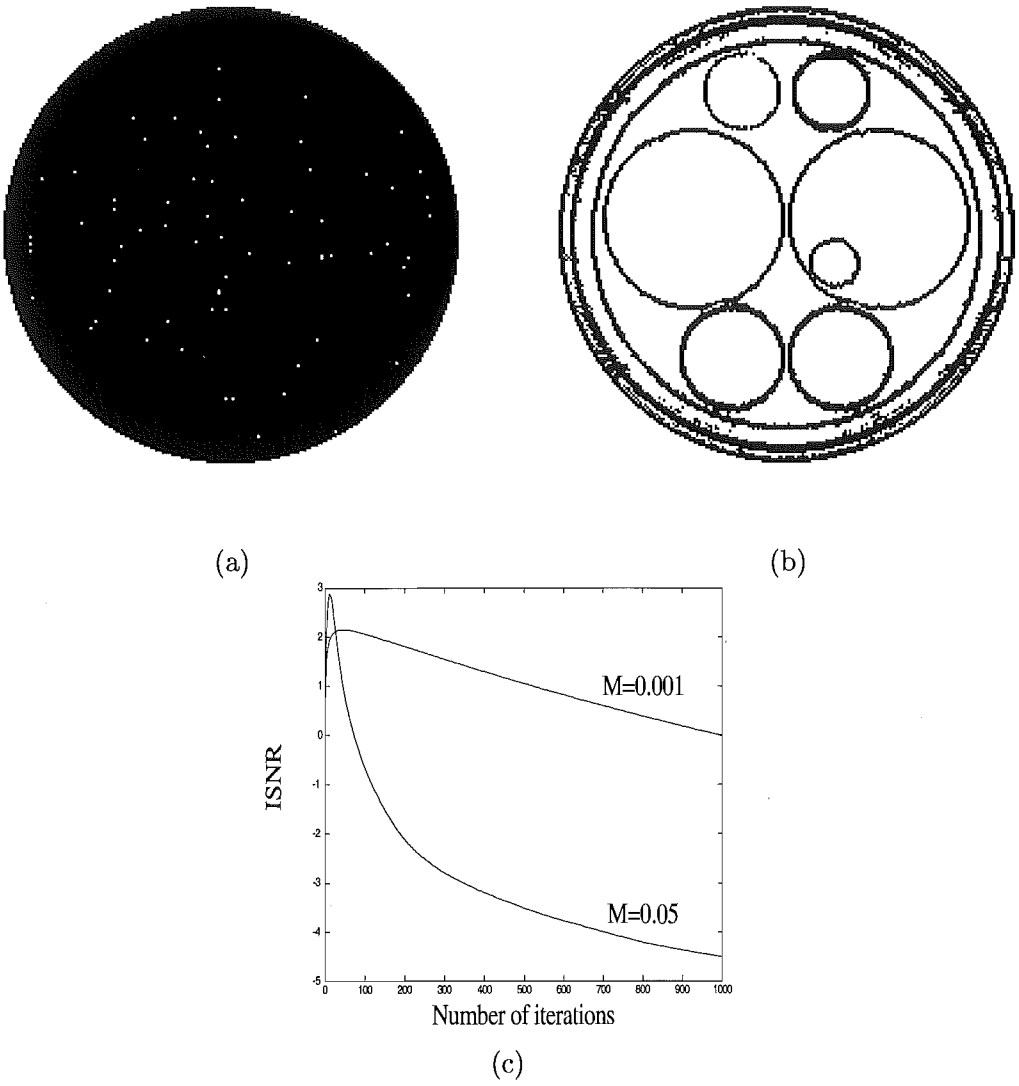


Figure 5.17 Truncated quadratic function: (a) regions below the threshold $M = 0.001$ shown in white; (b) regions below the threshold $M = 0.05$ shown in white; (c)ISNR curve obtained from applying the truncated quadratic function with $\lambda = 0.5$.

5.7 CONCLUSION

As discussed in section 5.1, the inverse problem of recovering f, given $Hf = G$, is ill-posed, and since the measured data G contains measurement errors, the exact solution,

if it exists, is useless. Therefore, regularization methods are required to transform an ill-posed problem into a well-posed problem. The most common regularization method is the Tikhonov regularization, where in general $\mathcal{S}(f) = \|Qf\|^2$, where Q is an operator. Common choices of Q are the identity operator, (known as the Tikhonov-Miller regularization) or $Q = \nabla$ (known as the smoothness constraint). Though both these methods do regularize the problem, they lack an edge preserving capability. For piecewise (blocky) images (e.g. the head image), the total variation method seems most effective since it regularizes the problem while preserving the edges as well. A compromise between smoothness and edge preservation enforcement can be obtained by applying other regularizing methods such as the regularization using the truncated quadratic function or the Huber function. It is important to note that nonconvex functions may be effective when applied on images with smooth patches and sharp boundaries, however convergence is not guaranteed.

Chapter 6

SUPERRESOLVING UNDERSAMPLED IMAGES

6.1 INTRODUCTION

A commonly used definition for 'superresolution', is the recovery of spatial frequency information beyond the limited bandwidth imposed by the imaging system. However, a broader definition of superresolution can include the recovery of missing spectral content inside the bandlimit [Hunt 1995]. In the conventional MR imaging technique (as discussed in chapter 2), transverse magnetization M_{xy} is sampled in a rectilinear fashion. An image is then reconstructed by applying the fast Fourier transform to the sampled collected data. The samples in the k-space are determined by the temporal evolution of the k value, i.e. the k-space trajectory which is specified by the time integral of the gradient (as shown in Eqn. 2.8). Since this technique allows considerable freedom in using different trajectories through the k-space, this opens up many possibilities for fast imaging techniques [Vlaardingerbroek and Boer 1996].

In MR imaging there is a continuous search for faster imaging methods to increase the throughput of the system, to freeze motion (to reduce motion artifacts) and for dynamic imaging. In addition to the fast imaging techniques (as discussed in section 2.3), imaging time can be reduced by sampling fewer points in the k-space according to some specific patterns. This is known as undersampling (discussed below) which will result in aliasing of the reconstructed image. The question therefore arises: can spectral samples be more widely spaced than the Nyquist limit and still enable the satisfactory recovery of a finite support image? Is the interpolation required to recover the full spectrum in this case easier to achieve than the extrapolation which corresponds to conventional superresolution?

Another application which encounters the undersampling problem is synthesis imaging. The aim of synthesis imaging is to obtain an image of the source from measurements of the visibility. The synthesis arrays sample the $u - v$ plane (spatial frequencies) at discrete locations. The distribution of sampled points in the $u - v$ plane is far from regular and typically contains a number of large gaps. Two techniques that are used in radioastronomy are CLEAN and maximum entropy methods

[Bracewell 1995, Thompson *et al.* 1986] which aim at setting the visibility at the unsampled points to values consistent with the most likely brightness distribution (Fourier transform of the visibility).

In this chapter different undersampling patterns are investigated and the recovery of the resultant aliased images is attempted. This chapter starts with a brief overview of undersampling and aliasing followed by a description of a direct method to recover an undersampled spectrum of a finite support object. In section 6.6 a description of the different undersampling patterns studied is presented followed by results obtained from the recovery of the undersampled images utilizing the iterative G-P algorithm.

6.2 UNDERSAMPLING AND ALIASING

As mentioned in section 2.5.3, in MR imaging the collected data is acquired by sampling the raw data $F(k_x, k_y)$ (the spectrum of an image $f(x, y)$) by a sampling function $S(k_x, k_y)$. The common and simplest sampling function is the infinite series of Dirac delta functions arranged in a rectangular grid, which will be used here to explain aliasing, where

$$S(p\Delta_{k_x}, q\Delta_{k_y}) = \frac{1}{\Delta_{k_x}\Delta_{k_y}} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \delta(k_x - p\Delta_{k_x}, k_y - q\Delta_{k_y}), \tag{6.1}$$

where Δ_{k_x} and Δ_{k_y} are the sampling intervals and $\frac{1}{\Delta_{k_x}}$ and $\frac{1}{\Delta_{k_y}}$ are the sampling rates along the k_x and k_y directions respectively. For brevity, the sampling function $S(p\Delta_{k_x}, q\Delta_{k_y})$ will be written as $S_{p,q}$. Since sampling occurs in the frequency domain, this corresponds to replication of the image $f(x, y)$ at intervals $\frac{1}{\Delta k_x}$ and $\frac{1}{\Delta k_y}$ in the x and y directions respectively. This is illustrated in Fig. 6.1, where the image $f(x, y)$ has a compact support of widths X and Y in the x and y direction respectively. Fig. 6.1 represents the case where the sampling rates are such that

$$\begin{aligned} \frac{1}{\Delta k_x} &> X \quad \text{and} \\ \frac{1}{\Delta k_y} &> Y, \end{aligned} \tag{6.2}$$

which is known as the oversampling case, and where the image $f(x, y)$ can be recovered completely. However if $F(k_x, k_y)$ is sampled at rates such that

$$\begin{aligned} \frac{1}{\Delta k_x} &< X \quad \text{and} \\ \frac{1}{\Delta k_y} &< Y, \end{aligned} \tag{6.3}$$

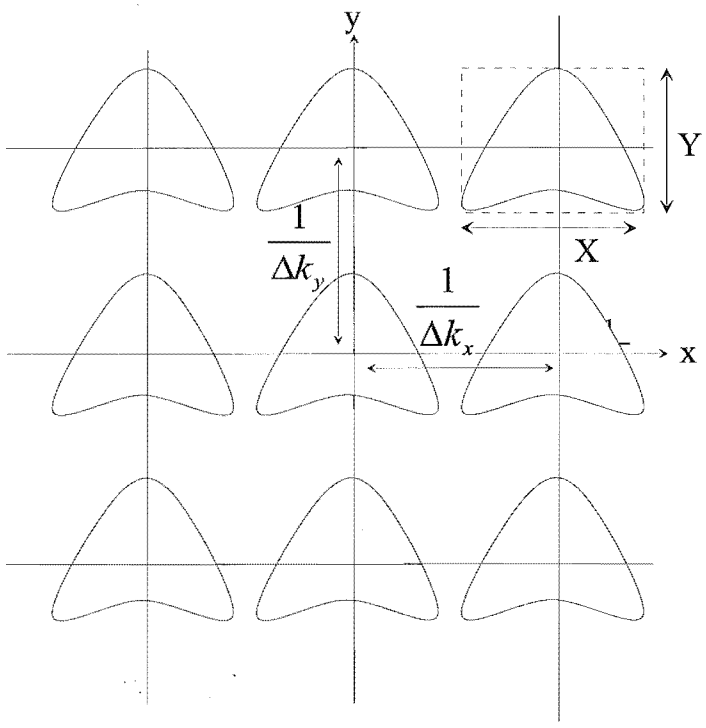


Figure 6.1 Replication of $f(x,y)$ in the space domain corresponding to the sampling of $F(k_x,k_y)$ in the frequency domain.

this is known as undersampling and aliasing of $f(x,y)$ will occur [Bracewell 1995], where pixels of the image outside $(\frac{1}{\Delta k_x},\frac{1}{\Delta k_y})$ will fold over to appear inside $(\frac{1}{\Delta k_x},\frac{1}{\Delta k_y})$ [Jain 1989]. If aliasing is not severe (i.e. the replications do not completely overlap each other) the resultant image will have a central zone not subjected to overlap but the remaining image will be the sum of the original image and the overlapping portions of the replications as shown in Fig. 6.2.

Since time is the limiting factor in MR imaging, imaging time can be reduced by undersampling the raw data, $F(k_x,k_y)$, according to a specific pattern. It is the aim of the work reported in the next sections to attempt to recover the missing samples by either applying a direct method or the G-P algorithm.

6.3 RECOVERY OF UNDERSAMPLED IMAGES

Up to this point all the superresolution methods have been applied on low resolution images obtained from bandlimiting a high resolution image. Applying the G-P algorithm on a bandlimited spectrum aims to recover spectral components beyond the imposed bandlimit as shown in Fig. 6.3(a) for the 1-D case. In this case the missing componenents are recovered by utilizing the information within the bandlimit which is a sort of extrapolation and therefore values of the recovered missing samples further from the bandlimit are likely to be more erroneous than those of the missing samples

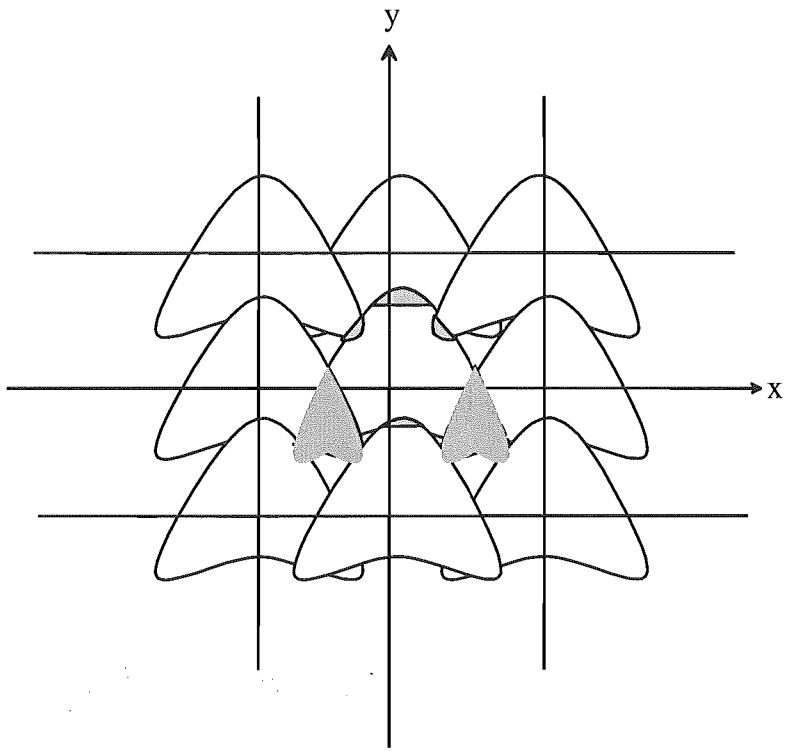


Figure 6.2 An illustration of aliasing.

closer to the bandlimit. However if the spectrum is undersampled as shown in Fig. 6.3(b), the missing components are recovered by utilizing information surrounding the missing sample, which represents a sort of interpolation.

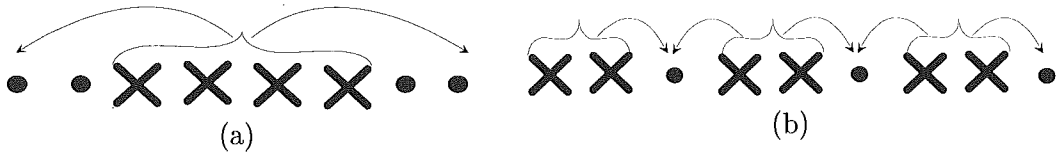


Figure 6.3 •: Missing samples, ×: sampled frequencies, (a) bandlimited case; (b) undersampled case.

In the next section a direct method is used to recover undersampled 1-D data and to study the effect of determinacy and the shape of the undersampling pattern on the amount of recovery achieved.

6.4 DIRECT METHOD TO RECOVER UNDERSAMPLED IMAGES

A direct method was presented by Walsh and Nielsen-Delaney [Walsh and Nielsen Delaney 1994] that supperresolves bandlimited images by taking advantage of the structure of the DFT and the prior knowledge of the support region of the image. In this

section the method is extended to the undersampled case as follows.

Assume a 1-D signal f of N samples $f_0, f_1, \ldots, f_{N-1}$ with a known support region of M samples where $M < N$ (i.e. $f_M = f_{M+1} \ldots = f_{N-1} = 0$) and with a spectrum F. F is only known at K frequencies such that $K < N$. Denote the set of measured spectral components by $\tilde{F}$ (with the unknown components set to zero) and the set of unknown spectral components by $\breve{F}$ (with the known components set to zero), therefore one can write

$$f_l = \frac{1}{N} \sum_{j=0}^{N-1} F_j w_N^{lj}, \quad \text{where } w_N = e^{\frac{i2\pi}{N}}, \quad \forall l \tag{6.4}$$

and

$$0 = \frac{1}{N} \sum_{j=0}^{N-1} \tilde{F}_j w_N^{lj} + \frac{1}{N} \sum_{j=0}^{N-1} \breve{F}_j w_N^{lj}, \quad l = M, M+1, \cdots, N-1, \tag{6.5}$$

Eqn. 6.5 represents a set of $N - M$ equations in $N - K$ unknowns and can be written as

$$A\underline{x} = \underline{b}, \quad \text{where}$$

$$A = \begin{bmatrix} w_N^{Mj_0} & w_N^{Mj_1} & \cdots & w_N^{Mj_{N-K-1}} \\ w_N^{(M+1)j_0} & w_N^{(M+1)j_1} & \cdots & w_N^{(M+1)j_{N-K-1}} \\ \vdots & \vdots & \ddots & \vdots \\ w_N^{(N-1)j_0} & w_N^{(N-1)j_1} & \cdots & w_N^{(N-1)j_{N-K-1}} \end{bmatrix} \tag{6.6}$$

$$\underline{x} = \begin{bmatrix} \breve{F}_{j_0} \\ \breve{F}_{j_1} \\ \vdots \\ \breve{F}_{j_{N-K-1}} \end{bmatrix} \quad \underline{b} = \begin{bmatrix} -\sum_{j=0}^{N-1} \tilde{F}_j w_N^{Mj} \\ -\sum_{j=0}^{N-1} \tilde{F}_j w_N^{(M+1)j} \\ \vdots \\ -\sum_{j=0}^{N-1} \tilde{F}_j w_N^{(N-1)j} \end{bmatrix}, \tag{6.7}$$

where $\breve{F}_{jk}$ represent the unknown spectral components, and the jk are the $N - K$ indices of the unknown spectral components.

Three cases need to be considered here: 1) $K = M$ resulting in a fully determined system where $\underline{x}$ can be solved by simply inverting A, 2) $K > M$ resulting in an over determined system which can be solved by using a least squares solution [Walsh and Nielsen Delaney 1994] such that $\underline{x} = (A^T A)^{-1} A^T \underline{b}$ or the singular value decomposition method (see section 8.2), and 3) when $K < M$ resulting in an underdetermined system, where the Moore-Penrose inverse can be used to obtain a minimum norm least squares solution [Sullivan and Liu 1984].

6.5 RESULTS

In this section, the direct method was applied on a 1-D signal in order to study the effect of the determinacy on the recovery process and to compare the amount of recovery between the bandlimited case and the undersampled case (i.e. to investigate if it is better to miss out a group of frequencies or miss out dispersed frequencies)

The three different cases of determinacy were studied by applying the direct method to a 1-D signal of 8 samples with a known support of 5 or 6 samples and 2 or 4 unknown spectral components. The results are shown in Table 6.1, where the first column shows the known support region of the 1-D signal, the second column shows the size of the support region (M), the third column shows the pattern of the spectral measurements, and the fourth column shows the number of measured spectral components (K). Depending on the the values of K and M for each case, the determinacy is shown in the fifth column. For each determinacy case, 2 different SNR cases were studied, 1) no noise (i.e. SNR=∞), 2) SNR=20dB as shown in the sixth column. The mean square error (MSE) between the original and recovered signal within the support region is used as an indicator of the amount of recovery achieved for the different cases. Note for the no noise over- and fully determined cases, the recovery is virtually perfect, and when the SNR=20dB, the increase in the MSE is associated with the addition of the noise. For the no noise underdetermined case (i.e. third row), the reconstruction is not as good as the previous two cases. The fourth row represents an underdetermined case but with the measured spectral components dispersed. The reconstruction in this case seems much better than in the underdetermined case when the measured spectral components are grouped together.

Support	M	Measurements ($\tilde{F}$)	K	Determinacy	SNR	MSE
[11111000]	5	$[F_0 0 F_2 F_3 F_4 F_5 F_6 0]$	6	over	∞	10^{-30}
					20dB	0.0029
[11111100]	6	$[F_0 0 F_2 F_3 F_4 F_5 F_6 0]$	6	fully	∞	10^{-30}
					20dB	0.0228
[11111000]	5	$[F_0 0 0 F_3 F_4 F_5 0 0]$	4	under	∞	0.0086
					20dB	0.0331
[11111000]	5	$[F_0 0 F_2 0 F_4 0 F_6 0]$	4	under	∞	0.0040
					20dB	0.0172

Table 6.1 Results obtained from applying the direct method on a 1-D signal for different cases of determinacy and noise.

The use of the direct method is limited to this section since it is difficult to apply it to problems with a substantial number of pixels. The iterative G-P algorithm is applied in the remainder of this chapter. Use of the direct method does however provide insight into the recovery process and may indicate the advantage of some sampling arrangements over others.

6.6 G-P ALGORITHM TO RECOVER UNDERSAMPLED MR IMAGES

In this section, different undersampling patterns are investigated by multiplying them by the spectrum of the high resolution image. The G-P algorithm with the positivity constraint (discussed in section 4.7.3) is then applied to the undersampled images. The different patterns studied are numbered 1 to 8 and are presented in subsections 6.6.1 to 6.6.8.

6.6.1 Pattern 1: $(S^1_{p,q})$

The first undersampling regular pattern $(S^1_{p,q})$ considered is shown in Fig. 6.4. The

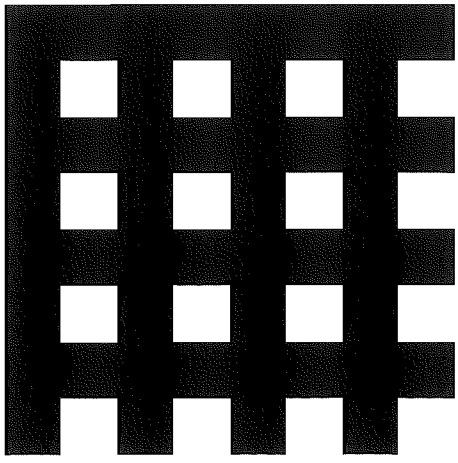


Figure 6.4 Undersampling pattern $S^1_{p,q}$, where a white pixel indicates a 1, and a black pixel indicates a 0.

sampled spectrum $\hat{F}_{p,q}$ results from multiplying the spectrum $F(k_x, k_y)$ by $S^1_{p,q}$ with the D.C. value set to zero. The resultant image is then obtained by utilizing the inverse discrete Fourier transform (IDFT) to produce the image $\hat{f}_{n,m}$. The relationship between the aliased image $\hat{f}_{n,m}$ and the imaged object $f_{n,m}$ can be found as follows.

$$S^1_{p,q} = \frac{1}{4}[1 + (-1)^{p+q} - (-1)^p - (-1)^q] \tag{6.8}$$

$$\hat{F}_{p,q} = F_{p,q} S^1_{p,q}. \tag{6.9}$$

Applying the inverse discrete Fourier transform on the sampled spectrum $\hat{F}_{p,q}$ we have

$$\begin{aligned} \hat{f}_{n,m} &= \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\hat{F}_{p,q}w_N^{np+mq} \qquad (6.10)\\ &= \frac{1}{4N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\hat{F}_{p,q}w_N^{np+mq}[1+(-1)^{p+q}-(-1)^p-(-1)^q] \end{aligned}$$

where $w_N = e^{\frac{i2\pi}{N}}$.

Since $(-1)^{kp+lq} = e^{i\pi(kp+lq)} = w_N^{\frac{N}{2}(kp+lq)}$, where k and l are constants, and from the properties of the discrete Fourier transform, we have

$$\hat{f}_{n+k,m+l} \quad \underset{\longleftrightarrow}{DFT} \quad F_{p,q}w_N^{(kp+lq)}, \qquad (6.11)$$

therefore

$$\hat{f}_{n,m} = \frac{1}{4}[f_{n,m}+f_{n-\frac{N}{2},m-\frac{N}{2}}-f_{n,m-\frac{N}{2}}-f_{n-\frac{N}{2},m}]. \qquad (6.12)$$

The undersampling pattern was applied on the high resolution noisy head data (where each of the spatial frequency components of the high resolution noise free head image, shown in Fig. 3.4(a), was corrupted by complex pseudorandom Gaussian noise so that the overall SNR was $30dB$), shown in Fig. 6.5(a). The resulting aliased image is shown in Fig. 6.5(b). Note the absence of Gibb's oscillation since no abrupt truncation was performed (as also obvious from Eqn. 6.12). The G-P algorithm with the positivity constraint was then applied to the aliased image resulting in the image shown in Fig. 6.5(c). In this case there is hardly any success in removing the aliasing within the support of the image, and the removal of the aliasing outside the support region is simply due to the application of the support constraint.

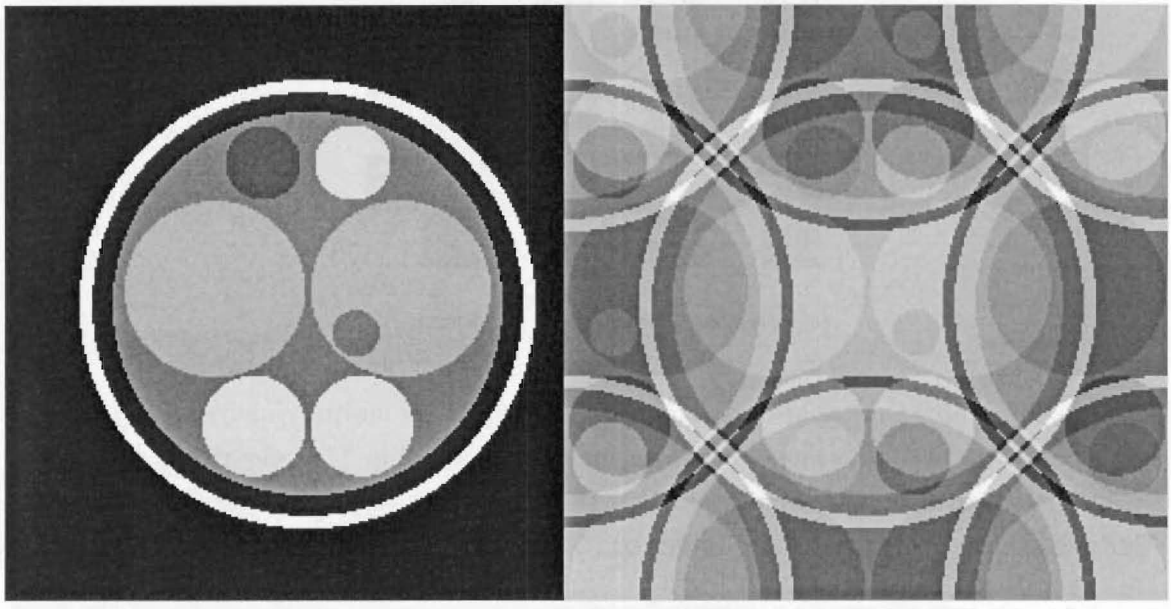
6.6.2 Pattern 2: $(S^2_{p,q})$

The second undersampling pattern $(S^2_{p,q})$ considered is shown in Fig. 6.6. In this case every other frequency of the spectrum of the high resolution image is missed out when multiplied by $S^2_{p,q}$, with the D.C. value being retained. The formulation for the aliased image $\hat{f}_{n,m}$ can be found as follows

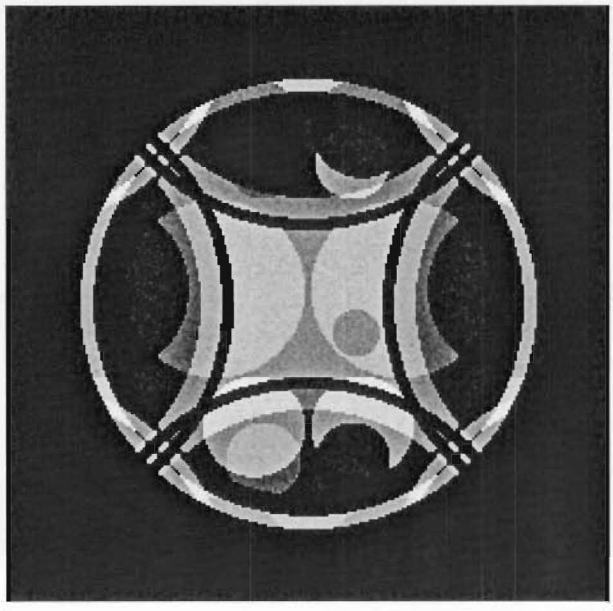
$$S^2_{p,q} = \frac{1}{2}[1+(-1)^{p+q}] \qquad (6.13)$$

$$\hat{F}_{p,q} = F_{p,q}S^2_{p,q}, \quad \text{therefore} \qquad (6.14)$$

$$\begin{aligned} \hat{f}_{n,m} &= \frac{1}{2N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\hat{F}_{p,q}w_N^{np+mq}[1+(-1)^{p+q}] \qquad (6.15)\\ &= \frac{1}{2}[f_{n,m}+f_{n-\frac{N}{2},m-\frac{N}{2}}]. \qquad (6.16) \end{aligned}$$



(a) (b)



(c)

Figure 6.5 (a) High resolution noisy head image; (b) undersampled head image using $S^1_{p,q}$; (c) recovered image utilizing G-P with the positivity constraint.

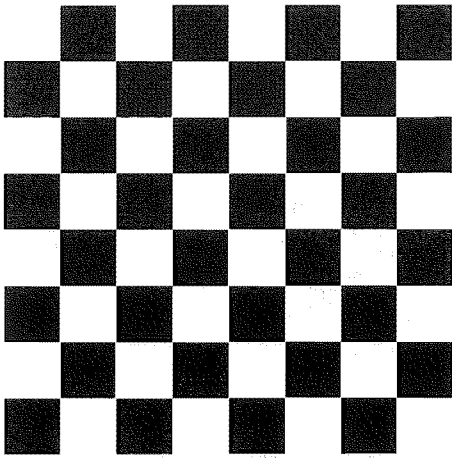
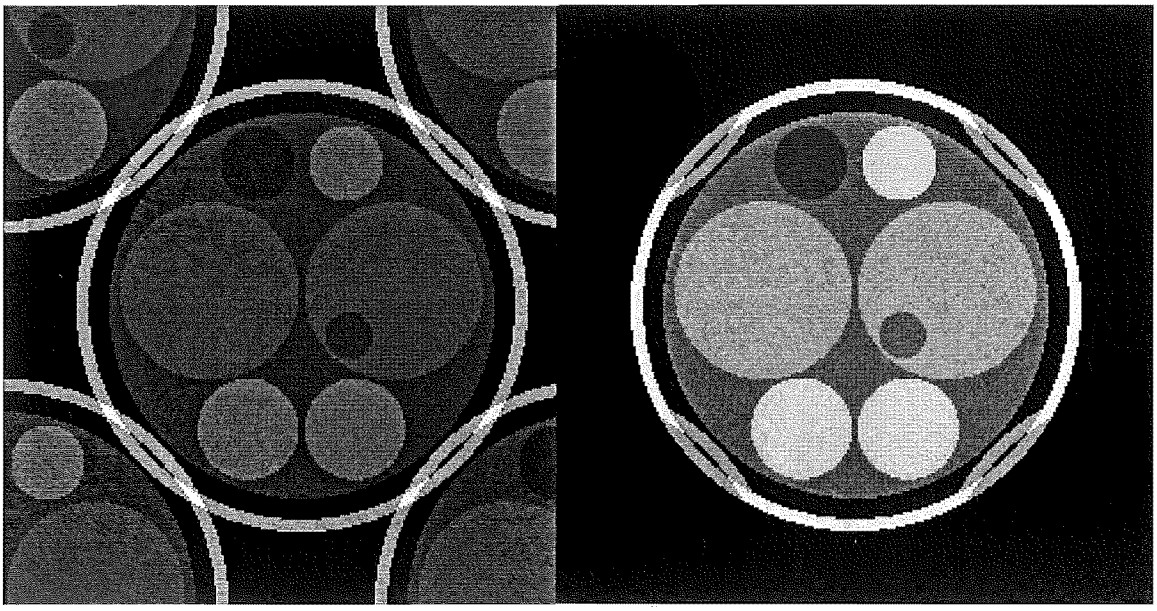


Figure 6.6 Undersampling pattern $S^2_{p,q}$.

The high resolution noisy head data was undersampled by multiplying its spectrum by $S^2_{n,m}$, and the resultant image is shown in Fig. 6.7(a). The G-P algorithm with the positivity constraint was then applied on the undersampled image, and the resultant image is shown in Fig. 6.7(b). Similar to the $S^1_{p,q}$ case most of the aliasing within the



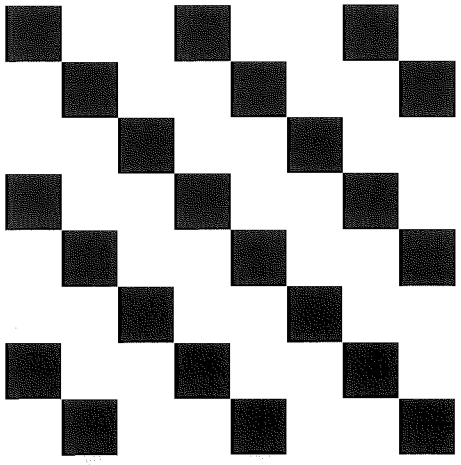
(a) (b)

Figure 6.7 (a) Undersampled head image using $S^2_{p,q}$; (b) recovered image utilizing G-P with the positivity constraint.

support region of the image was not removed.

6.6.3 Pattern 3: $(S^3_{p,q})$

Another undersampling pattern $(S^3_{p,q})$ shown in Fig. 6.8 was studied. The aliased



(a)

Figure 6.8 (a) Undersampling pattern $S^3_{p,q}$.

image resulting from the undersampling pattern $(S^3_{p,q})$ can be formulated as follows.

$$S^3_{p,q} = \frac{4}{3}\sin^2\frac{2\pi}{3}(p-q), \tag{6.17}$$

using the following trigonometric identity

$$\begin{aligned}
\sin^2 x &= \frac{1}{2}[1-\cos 2x] \quad \text{therefore} \\
S^3_{p,q} &= \frac{2}{3}[1-\cos\frac{4\pi}{3}(p-q)] \\
&= \frac{2}{3}-\frac{1}{3}[e^{\frac{i4\pi}{3}(p-q)}+e^{\frac{-i4\pi}{3}(p-q)}].
\end{aligned}$$

Applying the inverse discrete Fourier transform to $\hat{F}_{p,q} = F_{p,q}S^3_{p,q}$ yields

$$\begin{aligned}
\hat{f}_{n,m} &= \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1}\hat{F}_{p,q}w_N^{np+mq}[\frac{2}{3}-\frac{1}{3}[e^{\frac{i4\pi}{3}(p-q)}+e^{\frac{-i4\pi}{3}(p-q)}]] \qquad (6.18)\\
&= \frac{2}{3}f_{n,m}-\frac{1}{3}f_{n-\frac{N}{3},m-\frac{2N}{3}}-\frac{1}{3}f_{n-\frac{2N}{3},m-\frac{N}{3}}.
\end{aligned}$$

Fig. 6.9(b) shows the undersampled image resulting from multiplying the spectrum of the high resolution noisy brain image, shown in Fig. 6.9(a), by $S^3_{p,q}$ with the D.C. value missed out. The aliasing is quite severe and covers most of the image. The G-P algorithm with the positivity constraint was applied on the undersampled image. The recovered image is shown in Fig. 6.9(c). In this case the G-P algorithm was capable of recovering most of the image and clearing most of the aliasing.

6.6.4 Pattern 4: $(S^4_{p,q})$

The fourth undersampling pattern studied is shown in Fig. 6.10. The aliased image

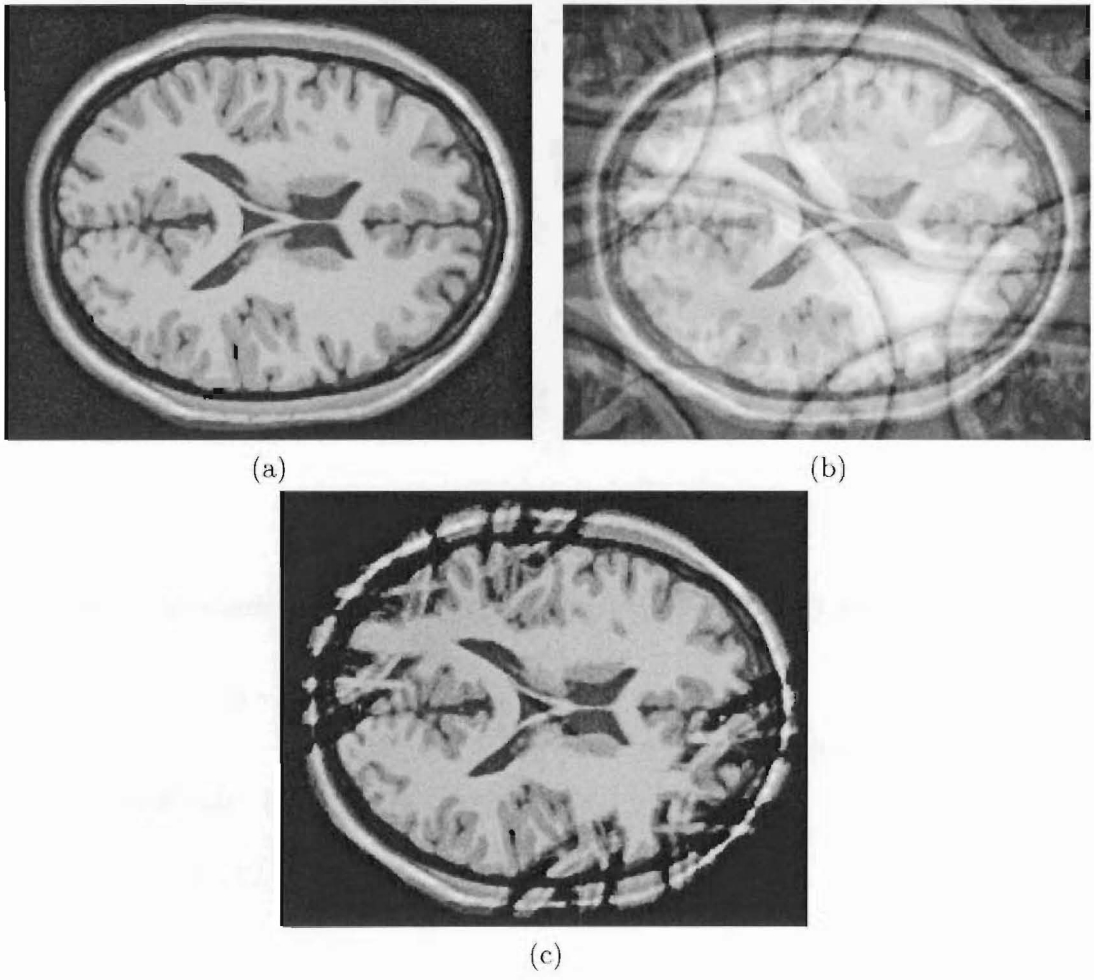


Figure 6.9 (a) High resolution noisy brain image; (b) undersampled brain image using $S^3_{p,q}$; (c) recovered image utilizing G-P with the positivity constraint.

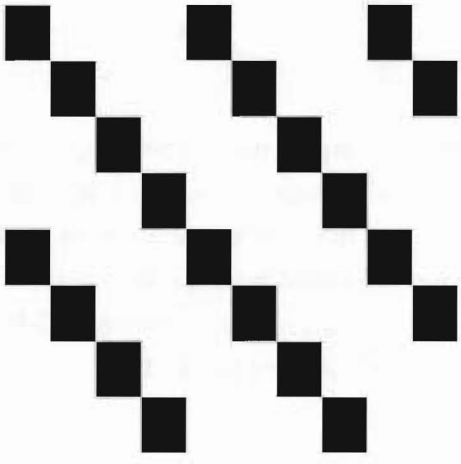


Figure 6.10 Undersampling pattern $S^4_{p,q}$.

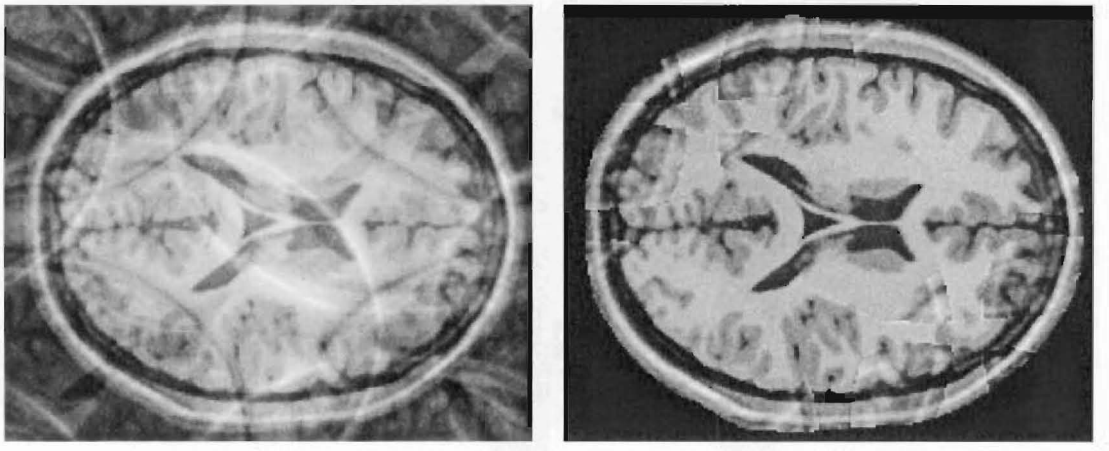
resulting from multiplying the spectrum $F_{p,q}$ by $S^4_{p,q}$ can be formulated as follows.

$$\begin{aligned} S^4_{p,q} &= \frac{1}{2}[1 + \cos\frac{\pi}{2}(p-q-2) + \sin^2\frac{\pi}{2}(p-q)] \\ &= \frac{1}{2}[\frac{3}{2} - \cos\frac{\pi}{2}(p-q) - \frac{1}{2}\cos\pi(p-q)] \\ &= \frac{3}{4} - \frac{1}{4}[e^{i\frac{\pi}{2}(p-q)} + e^{-i\frac{\pi}{2}(p-q)}] - \frac{1}{8}[e^{i\pi(p-q)} + e^{-i\pi(p-q)}]. \end{aligned} \tag{6.19}$$

Multiplying the spectrum $F_{p,q}$ by $S^4_{p,q}$ and applying the inverse discrete Fourier transform gives

$$\hat{f}_{n,m} = \frac{3}{4}f_{n,m} - \frac{1}{4}[f_{n-\frac{3N}{4},m-\frac{N}{4}} + f_{n-\frac{N}{4},m-\frac{3N}{4}}] - \frac{1}{4}f_{n-\frac{N}{2},m-\frac{N}{2}}. \tag{6.20}$$

The undersampling pattern $S^4_{p,q}$ was applied to the high resolution noisy brain data with the D.C. value retained and the resulting aliased image is shown in Fig. 6.11(a). The G-P algorithm with the positivity constraint was applied on the aliased image resulting in the recovered image shown in Fig. 6.11(b), where most of the aliasing within the support region of the image was removed.



(a) (b)

Figure 6.11 (a) Undersampled brain image using $S^4_{p,q}$; (b) recovered image utilizing G-P with the positivity constraint.

It is important to mention at this point that if the conventional rectilinear k-space sampling method, discussed in section 2.2.2, was used to generate the undersampling patterns $S^2_{p,q}$, $S^3_{p,q}$ and $S^4_{p,q}$, no reduction in imaging time would be achieved. This is because to generate those patterns all the phase encoding steps must be performed. However if the undersampling patterns were rotated such that the missing data were along the frequency encoding direction, then specific phase encoding steps (corresponding to the missing data) are not required, thereby reducing imaging time. It is important to note that this would not affect the reconstruction process since a rotation in the k-space

domain results in a similar rotation in the image domain, according to the following equation which utilizes the polar form of the two dimensional Fourier transform

$$\begin{aligned} f(r,\phi) &= \int_0^{2\pi}\int_0^{\infty} F(k,\theta+\theta_0)e^{i2\pi rk\cos(\theta-\phi)}\ k\ dk\ d\theta, \\ &= \int_0^{2\pi}\int_0^{\infty} F(k,\theta')e^{i2\pi rk\cos(\theta-(\phi+\theta_0))}\ k\ dk\ d\theta', \\ &= f(r,\phi+\theta_0), \end{aligned} \tag{6.21}$$

where $x = r\ cos(\phi), y = r\ sin(\phi), k_x = k\ cos(\theta), k_y = k\ sin(\theta), \theta' = \theta + \theta_0$, F is the collected k-space data and f is the corresponding image.

6.6.5 Pattern 5: $(S^5_{p,q})$

The next undersampling pattern studied is shown in Fig. 6.12 where it is expressed mathematically as

$$S^5_{p,q} = \frac{1}{4}(\frac{3}{2} - \cos(\frac{\pi}{2}p) - \frac{1}{2}\cos(\pi p))(\frac{3}{2} - \cos(\frac{\pi}{2}q) - \frac{1}{2}\cos(\pi q)). \tag{6.22}$$

The aliased image resulting from multiplying the spectrum $F_{p,q}$ by $S^5_{p,q}$ can be formu-

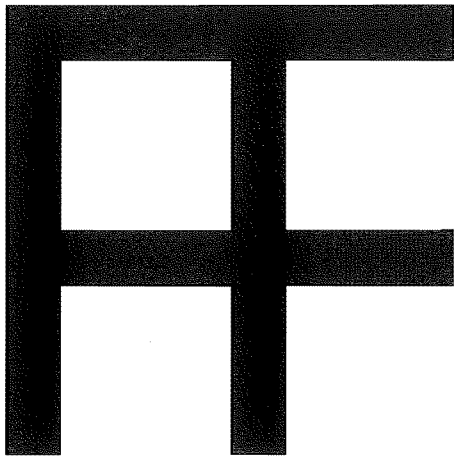
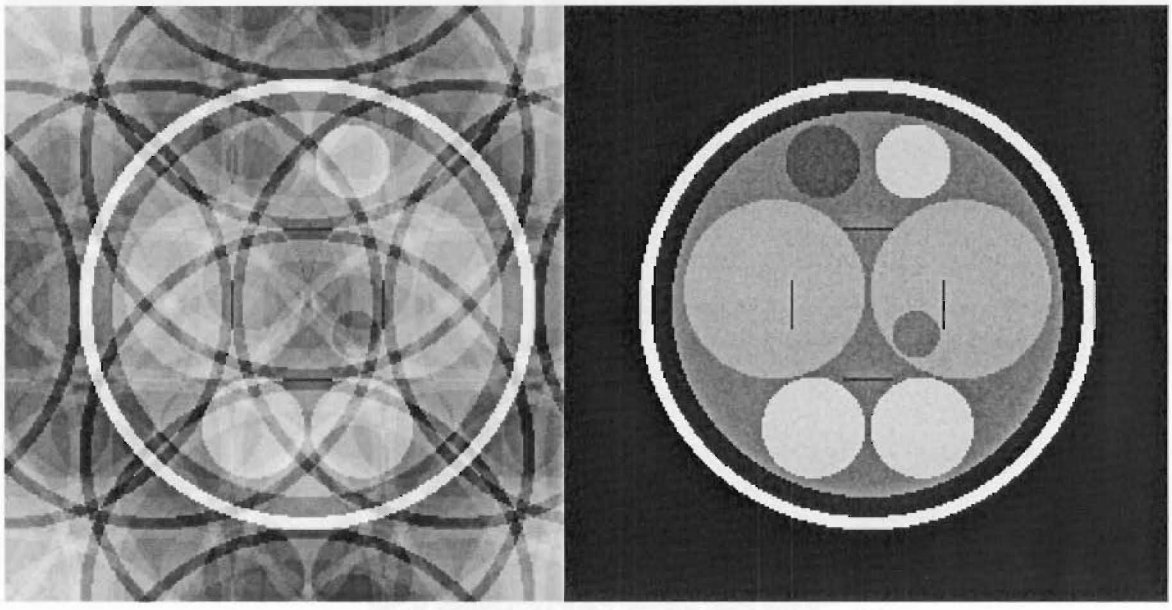


Figure 6.12 Undersampling pattern $S^5_{p,q}$.

lated as follows.

$$\begin{aligned} \hat{f}_{n,m} = \frac{1}{16}[&f_{n,m} - 3f_{n,m-\frac{3N}{4}} - 3f_{n,m-\frac{N}{4}} - 3f_{n,m-\frac{N}{2}} \\ &-3f_{n-\frac{3N}{4},m} - 3f_{n-\frac{N}{4},m} - 3f_{n-\frac{N}{2},m} \\ &+f_{n-\frac{3N}{4},m-\frac{3N}{4}} + f_{n-\frac{3N}{4},m-\frac{N}{4}} + f_{n-\frac{N}{4},m-\frac{3N}{4}} \\ &+f_{n-\frac{N}{4},m-\frac{N}{4}} + f_{n-\frac{3N}{4},m-\frac{N}{2}} + f_{n-\frac{N}{4},m-\frac{N}{2}} \\ &+f_{n-\frac{N}{2},m-\frac{3N}{4}} + f_{n-\frac{N}{2},m-\frac{N}{4}} + f_{n-\frac{N}{2},m-\frac{N}{2}}]. \end{aligned} \tag{6.23}$$

The undersampling pattern ($S^5_{p,q}$), shown in Fig. 6.12, was multiplied by the spectrum of the high resolution noisy head data (missing out the D.C. value). The aliasing in the resultant undersampled image shown in Fig. 6.13(a) seems quite severe and fully overlaps the original image. The G-P algorithm with the positivity constraint was then applied on the aliased image resulting in the recovered image shown in Fig. 6.13(b), which shows an unexpected nearly complete recovery of the image. The undersampling pattern $S^5_{p,q}$ was also applied on the high resolution brain image, however the recovery of the undersampled image was not as successful as it was for the head image. The difference is because of the more extensive separation between the support boundary and the image frame in the case of the head image. This is further discussed in section 6.7.



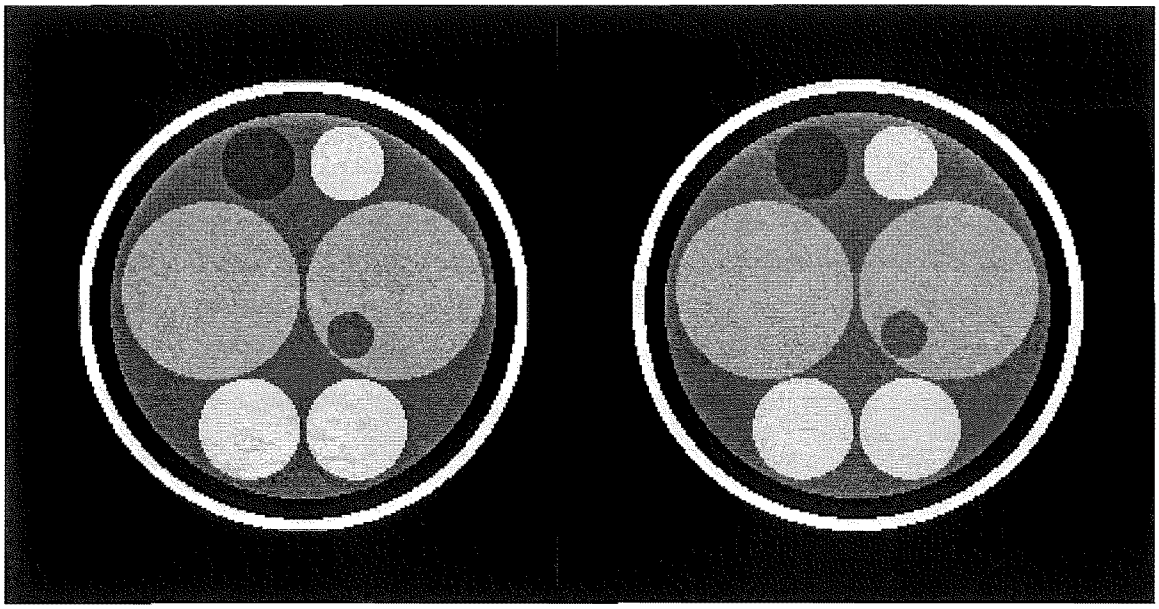
(a) (b)

Figure 6.13 (a)Undersampled head image using ($S^5_{p,q}$); (b) recovered image utilizing G-P with the positivity constraint.

Since the artifacts in the recovered image shown in 6.13(b) are of small widths and lie within a region of uniform value, the neighbourhood based quantization method discussed in subsection 4.9.2 was applied on the recovered image and the resultant image is shown in Fig. 6.14(a). In this case there was a complete elimination of the artifact.

Another way of removing the artifact in the recovered image, shown in Fig. 6.13(b), is by applying the total variation regularization method with $\lambda = 0.05$ and $\alpha^k = 0.05$ and the recovered image is shown in Fig. 6.14(b), where a complete elimination of the aliasing was achieved. The reason behind using the total variation method is that the black horizontal and vertical lines (artifacts) in the recovered image of Fig. 6.13(b), represent a relatively large jump for a very short length (i.e. a large total variation).

Therefore the method of total variation (which aims at minimizing the total variation) is suitable for removing this artifact.



(a) (b)

Figure 6.14 (a) Recovered image obtained from applying neighbourhood based quantization; (b) recovered image applying total variation with $\lambda = 0.05$ and $\alpha^k = 0.05$.

6.6.6 Pattern 6: $(S^6_{p,q})$

The next undersampling pattern studied is shown in Fig. 6.15. The undersampled

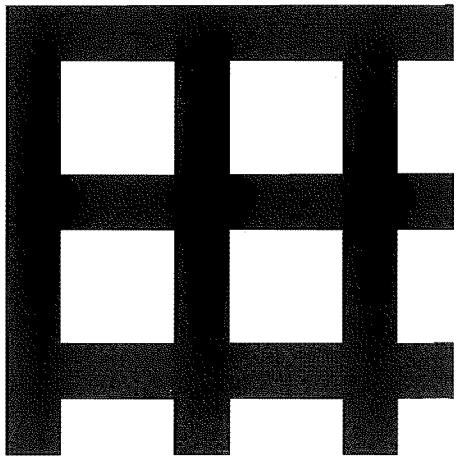


Figure 6.15 Undersampling pattern $S^6_{p,q}$.

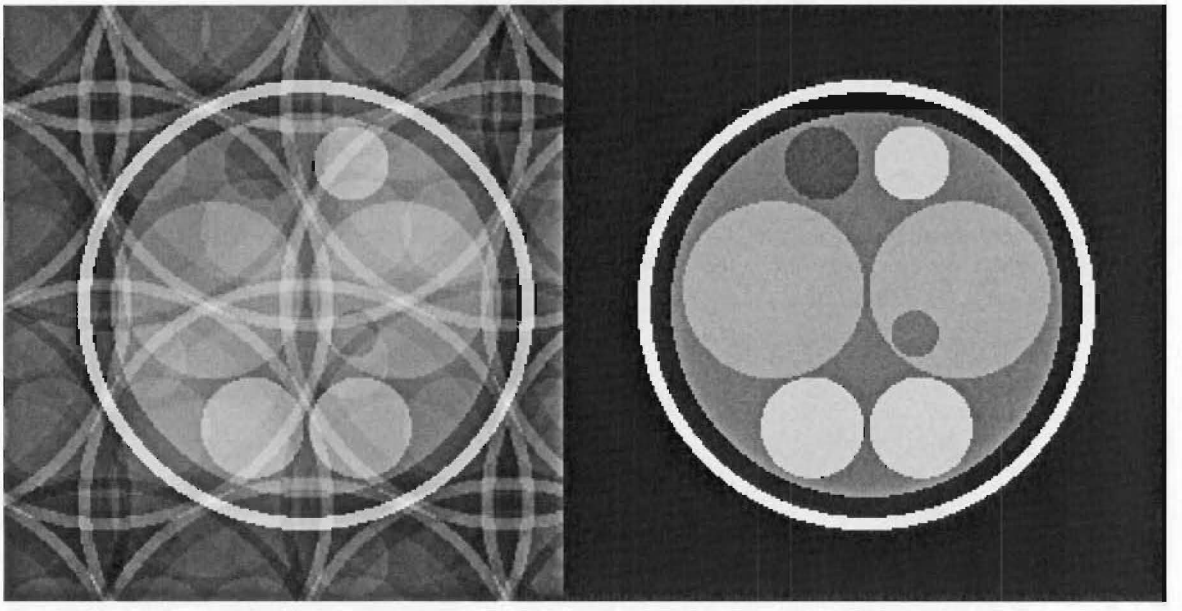
image using this pattern can be formulated as follows.

$$S^6_{p,q} = \frac{-32}{9}[\sin^2(\frac{2\pi p}{3})\sin^2(\frac{2\pi q}{3})\cos(\frac{2\pi(p+q)}{3})\cos(\frac{2\pi(p-q)}{3}). \quad (6.24)$$

Multiplying the spectrum $F_{p,q}$ by $S^6_{p,q}$ and applying the inverse discrete Fourier transform gives

$$\begin{aligned}\hat{f}_{n,m} &= -\frac{1}{18}[4f_{n-\frac{2N}{3},m} + 4f_{n,m-\frac{2N}{3}} + 4f_{n-\frac{N}{3},m} + 4f_{n,m-\frac{N}{3}} \\ &\quad -f_{n-\frac{2N}{3},m-\frac{2N}{3}} - 2f_{n-\frac{N}{3},m-\frac{N}{3}} - 2f_{n-\frac{2N}{3},m-\frac{N}{3}} - 3f_{n-\frac{N}{3},m-\frac{2N}{3}} \\ &\quad -8f_{n,m}] \end{aligned} \tag{6.25}$$

This undersampling pattern was applied to the high resolution noisy head image (with D.C. value missed out) resulting in the undersampled image shown in Fig. 6.16(a). The G-P algorithm with the positivity constraint was applied on the aliased image and the resultant image is shown in Fig. 6.16(b). Again, even though the aliasing seems severe and covers the whole image, the G-P algorithm was capable of completely removing the aliasing. The undersampling pattern $S^6_{p,q}$ was also applied on the high resolution brain image, however the recovery of the undersampled image was not as successful as the result obtained for the head image.



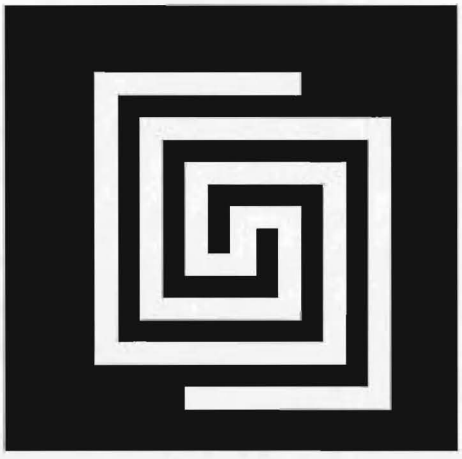
(a) (b)

Figure 6.16 (a)Undersampled head image using $(S^6_{p,q})$; (b) recovered image utilizing G-P with the positivity constraint.

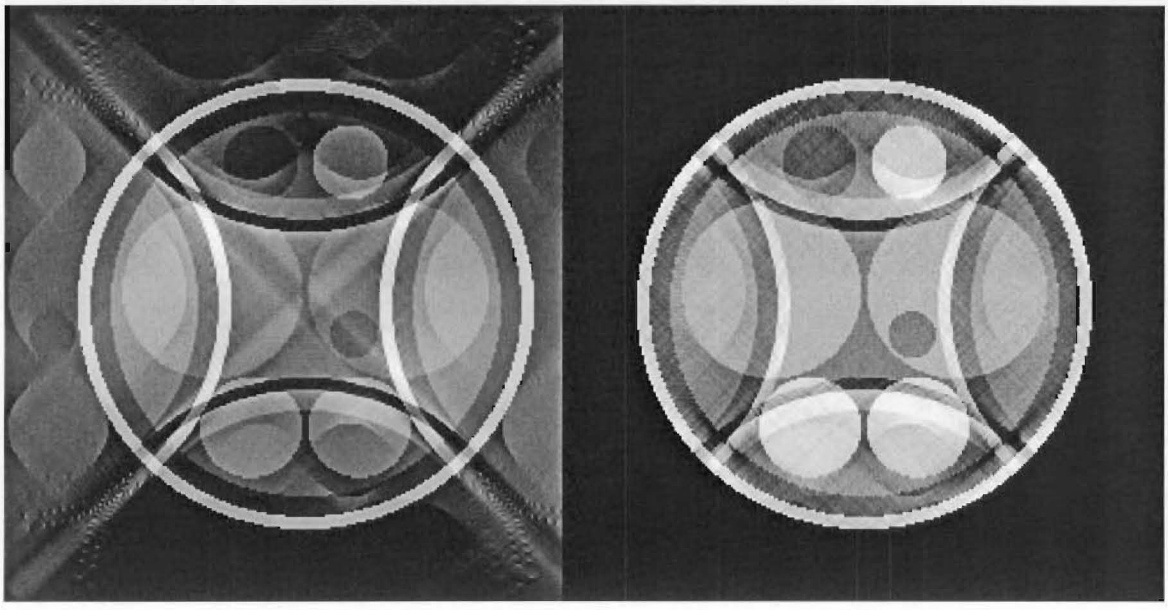
6.6.7 Pattern 7: $(S^7_{p,q})$

Until now, the patterns studied were obtained by a regular subsampling of the grid. A more practical, but irregular, undersampling pattern, shown in Fig. 6.17(a), was also studied. As mentioned in 2.3.2, spiral scans are more efficient and faster than rectilinear scans [Hennig 1999], therefore by utilizing spiral undersampling patterns, imaging time

can be reduced even further. The undersampling pattern, $S^7_{p,q}$, was applied to the high resolution head image (retaining the D.C. value) resulting in the aliased image shown in Fig. 6.17(b). The G-P algorithm was then applied on the undersampled image, and the recovered image is shown in Fig. 6.17(c), where some of the aliasing within the support region was removed.



(a)



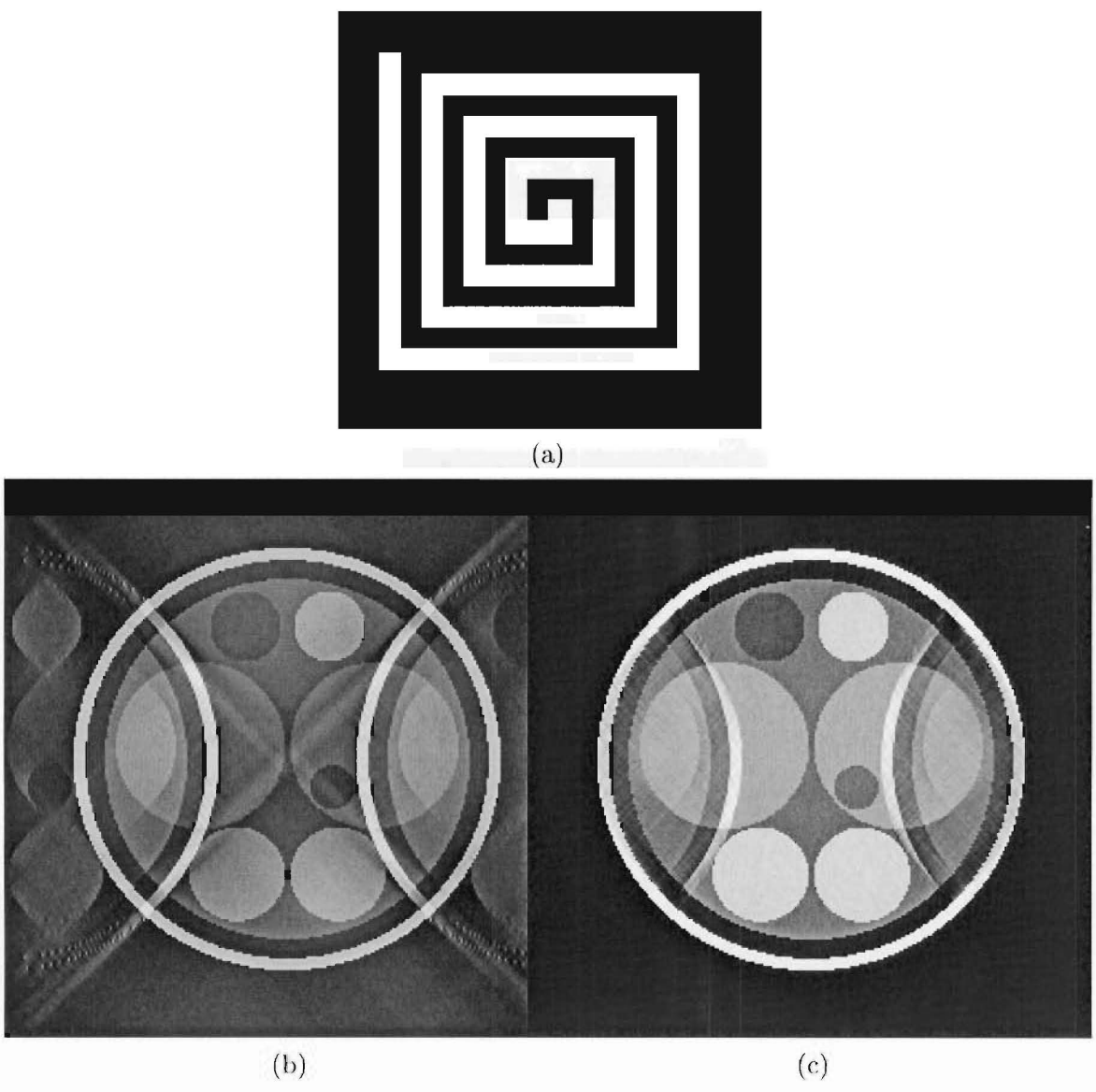
(b) (c)

Figure 6.17 (a)Undersampling pattern $S^7_{p,q}$; (b)undersampled head image; (c) recovered image utilizing the G-P algorithm.

6.6.8 Pattern 8: $(S^8_{p,q})$

Another possible spiral pattern studied is shown in Fig. 6.18(a). This pattern was applied to both the high resolution noisy head and brain data, and the resultant corresponding undersampled images are shown in Fig. 6.18(b) and 6.19(b) respectively.

Fig. 6.18(c) and 6.19(c) show the corresponding recovered images obtained from applying the G-P algorithm on the undersampled images. Again, in both cases, some of the aliasing within the support region was removed. Note in this case the Hermitian property is not satisfied.



(a)

(b) (c)

Figure 6.18 (a)Undersampling pattern $S^7_{p,q}$; (b)undersampled head image; (c) recovered image utilizing the G-P algorithm.

Since the Hermitian property is not satisfied in this case, it would be very advantageous to utilize the conjugate symmetry property discussed in section 2.3.4 to fill in the missing data, which can be expected to significantly improve the recovery process.

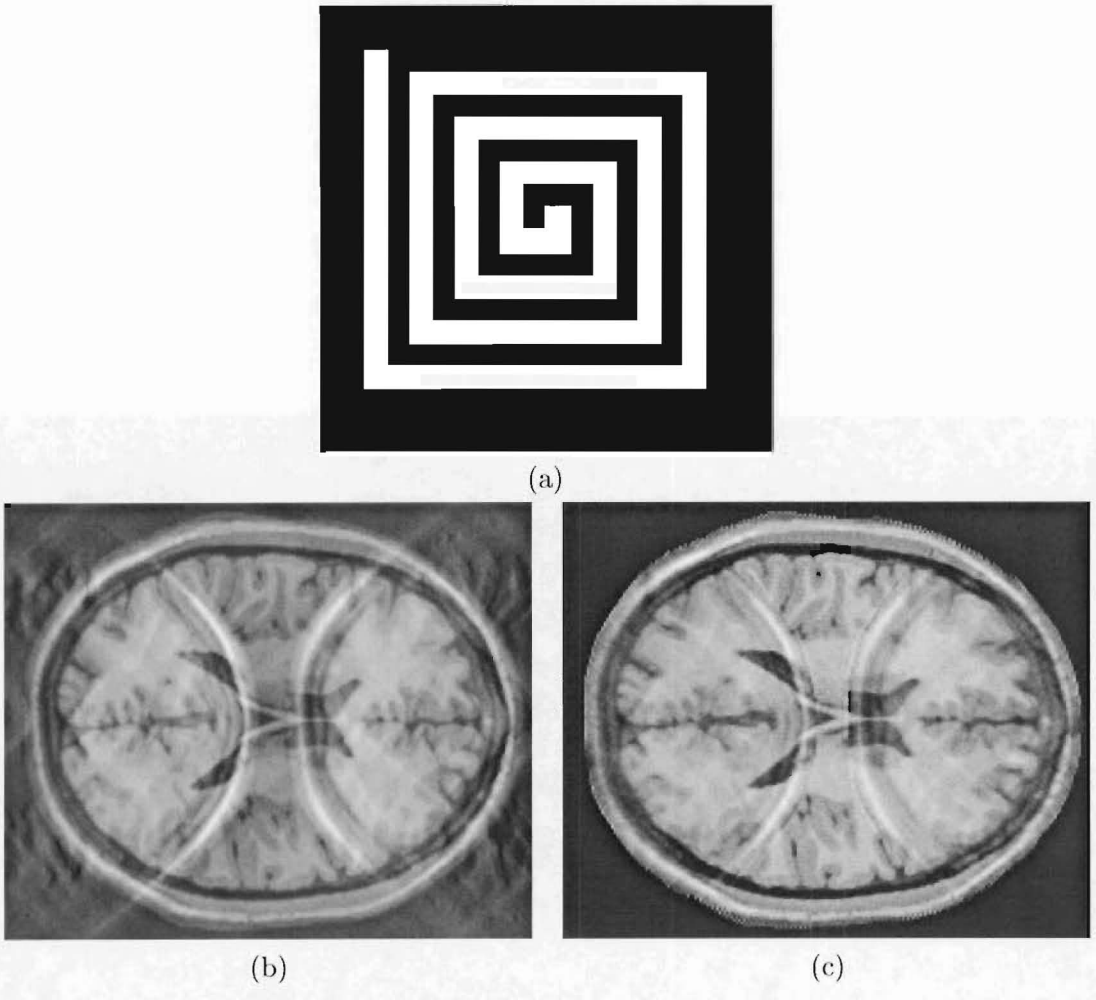


Figure 6.19 (a)Undersampling pattern $S^8_{p,q}$; (b)undersampled head image; (c) recovered image utilizing the G-P algorithm.

6.7 ANALYSIS OF RESULTS

This section attempts to explain some of the results obtained in section 6.6 for the different undersampling patterns used. Fig. 6.20 is a summary of the six regular undersampling patterns used, the undersampled or blurred head and brain images and their corresponding recovered images obtained from applying the G-P algorithm with the positivity constraint. The last columns list the ratios (for the head and brain images) of the number of missing spectral components to the number of pixels outside the support region (which are known to be zero) and is denoted by $\mathcal{R}$. When $\mathcal{R} \leq 1$, this means that the number of missing components is less than or equal to the number of pixels outside the support region resulting in a fully or overdetermined system (since the number of unknowns is less than or equal to the number of knowns).

In this section, the dependence of the amount of recovery achieved on both $\mathcal{R}$ and the shape of the undersampling pattern is discussed. The more the information (i.e. blocks of retained spectral components) surrounding a missed out component resulting from the application of an undersampling pattern, the better the recovery.

In order to illustrate this point, consider patterns $S^1_{p,q}$ and $S^2_{p,q}$. In both cases there is no recovery of the aliased region in the head images. With regards to $S^1_{p,q}$ this is due to the fact that $\mathcal{R} = 1.4$ indicating an underdetermined system. However, for $S^2_{p,q}$, even though $\mathcal{R} = 0.9$ indicating an overdetermined system, the shape of the undersampling pattern seems to affect the recovery, since each missed out frequency is surrounded by equal numbers of retained and missed out frequencies. With regards to the head image, the above two conditions seem to comply with results obtained for the remaining undersampling patterns.

For the brain image, both $S^1_{p,q}$ and $S^2_{p,q}$ result in an underdetermined system, which explains the absence of recovery of the aliased regions. Even though pattern $S^3_{p,q}$ also results in an underdetermined system, significant recovery of the aliased regions is due to the specific pattern used where most missed out frequencies are surrounded by two retained frequencies in most directions. Improved results are obtained when pattern $S^4_{p,q}$ is used, since that results in an overdetermined system and most missed out frequencies are surrounded by three retained frequencies in most directions. Since both patterns $S^5_{p,q}$ and $S^6_{p,q}$ result in underdetermined systems, hardly any recovery of the aliased regions is achieved. In addition, the brain image seems extremely sensitive to the shape of the undersampling patterns $S^5_{p,q}$ and $S^6_{p,q}$, where many missed out frequencies are surrounded by blocks of missed out frequencies. It is also important to mention that in all the above undersampling patterns, the ISNR curves for both sets of data were levelled off and most of them reached the maximum in less than 10 iterations.

Since there seems to be a certain pattern in the recovery process, it might be useful in this case to compare the values of the restored frequencies to their correct values. A

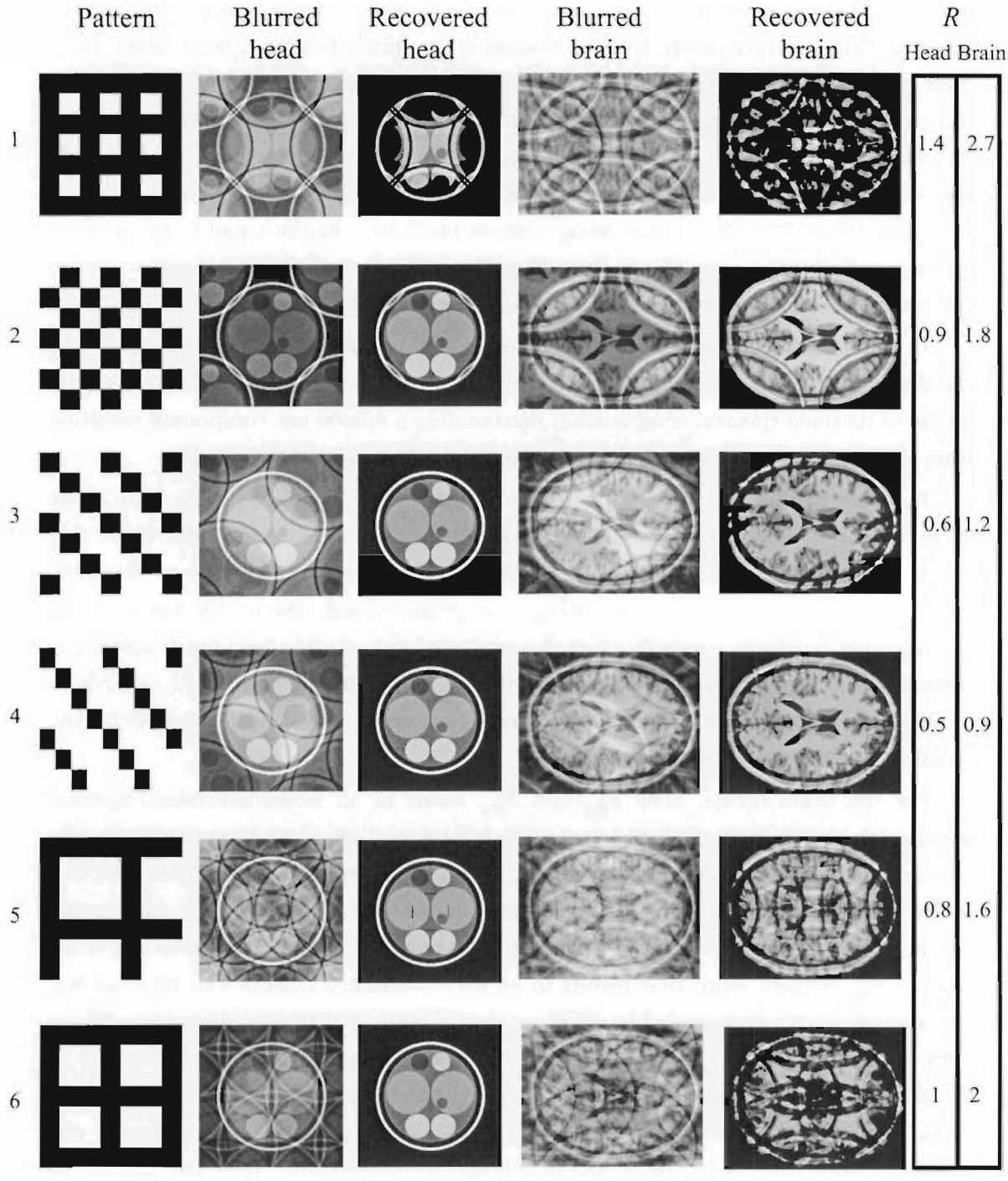


Figure 6.20 List of undersampling patterns, aliased images and the recovered images for both head and brain images.

pattern in the restoration of frequencies might be noticed which may help explain why some regions in the aliased images are completely recovered and others are unaffected.

6.8 CONCLUSION

In this chapter instead of the conventional reduction in imaging time by bandlimiting an image, undersampled images were studied. Different undersampling patterns were investigated by applying them on both the high resolution head and brain data. For some undersampling patterns, significant recovery (and in some cases complete recovery) was achieved by using the simple G-P algorithm with the positivity constraint even though in many cases aliasing seemed quite severe. The use of additional *a priori* information when appropriate was also found to improve the reconstruction achieved.

When bandlimiting an image, blocks of information are lost resulting in a loss of fine details in the image and superresolution aims to recover this loss. However when an image is undersampled, parts of the image simply overlap other parts, which is known as aliasing (all the information is there but simply superimposed). This seems to be an easier problem for the superresolution algorithms to solve.

Chapter 7

SUPERRESOLVING A REGION OF INTEREST

7.1 INTRODUCTION

In applications where time is a limiting factor, for example in cases where the patient is a child or is a claustrophobic [PHILIPS], Gibb's oscillations and loss of spatial resolution may degrade the image significantly. Dynamic imaging is such an application where a time series of images of a specific slice or volume are required over a short period of time. In most dynamic imaging applications, intensities of specific regions vary with time, while the main anatomical structures remain stationary. Examples are real time monitoring of the movement of organs and the intravenous injection of a contrast medium to study a specific organ. The difference between the different dynamic slices is usually small and localized to a specific region. Thus when monitoring the insertion of a biopsy needle, only the regions adjacent to the path of the needle will vary. In case of a tumour in the brain being followed over a course of treatment, only the tumour region is expected to vary significantly from one dynamic image to the next. If in each dynamic imaging session, high resolution images (i.e. complete data sets) are acquired, there would be substantial redundancy. Several methods have been proposed to attempt to utilize the particular attributes of dynamic imaging in order to reduce the scanning time. These are reviewed and additional methods are suggested in this chapter.

The chapter starts with a study of some of the techniques used in reducing imaging time, in particular the reduced acquisition, keyhole, and reduced field of view methods, and results are presented for each of these. Section 7.5 studies a new technique that can be applied to dynamic imaging, and results are presented to be compared with the existing methods.

7.2 REDUCED ACQUISITION METHOD

Since imaging time is proportional to the amount of collected data, the most common method of reducing imaging time is reducing the amount of acquired data. This can be achieved by acquiring the central portion of the k-space only and then setting the

remaining spectral components to zero. This is known as zero filling and is represented in Fig. 7.1. In many applications a reduced acquisition of 80% of the full data set is applied. This results in a low resolution blurred image which suffers from Gibb's oscillation. Nevertheless, this technique is still useful in cases where fine details are not that important and reducing imaging time is a priority.

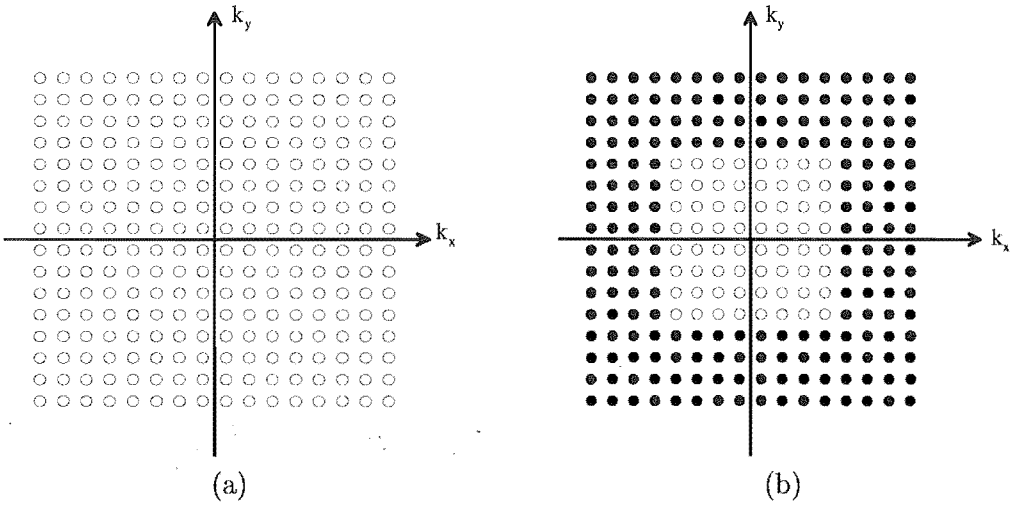
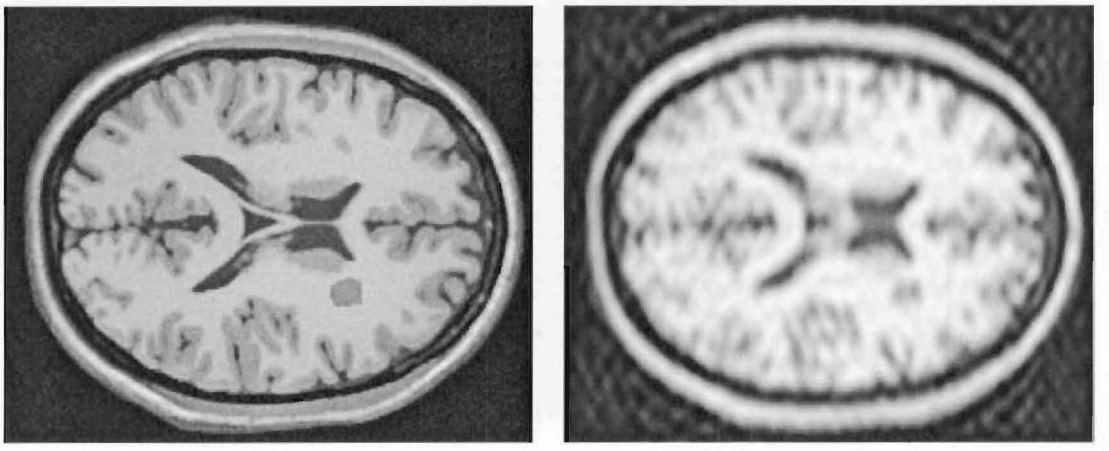


Figure 7.1 (a) A complete k-space data set; (b) a reduced k-space data set, where ○ represents a collected datum and • represents a missing datum which is set to zero.

A high resolution noisy brain image with a 'tumour' was created by adding a tumour of arbitrary shape to the high resolution noise free image and then each spatial frequency component was corrupted by complex pseudorandom Gaussian noise so that the overall SNR was $30dB$. Fig. 7.2(a) shows the inverse DFT of the corrupted spectrum, which is referred to as the tumour image. The reduced acquisition method was applied on the spectrum of the high resolution tumour image, where only the lowest 31 spatial frequencies vertically and the lowest 71 spatial frequencies horizontally were collected and the remaining missing components were set to zero. The inverse DFT was then applied on the modified spectrum resulting in the image shown in Fig. 7.2(b) and is referred to as the bandlimited or low resolution tumour image. The image obtained from this method is very blurred and suffers from Gibb's oscillation due to the abrupt truncation of the high frequency data. While the tumour is visible it has clearly not been imaged precisely. Note that, its size appears reduced in the low resolution image (Fig. 7.2(b))

7.3 KEYHOLE METHOD

As mentioned above, the zero filling method results in ringing artifacts and blurring. A method known as the keyhole (or the k-substitution) method, presented independently by Van Vaals and Jones [van Vaals *et al.* 1993, Jones *et al.* 1993], takes advantage of



(a) (b)

Figure 7.2 Image of brain with 'tumour': (a) high resolution noisy image; (b) low resolution image.

the fact that most of the image information (i.e. the coarse appearance of the image) lies in the low spatial frequency region whereas the finer details (e.g. sharpness of edges) lie in the high spatial frequency region. In this method a reference or base image is scanned, where a complete data set is collected as shown in Fig. 7.3(a), before performing a dynamic study. During the dynamic study only the central region of the k-space (the 'keyhole') is acquired as shown in Fig. 7.3(b), which is then substituted for the corresponding central region of the spectrum of the reference image as shown in Fig. 7.3(c). The new spectrum is therefore a composite of low frequency data from the dynamic (fast) study with high frequency data from the original (slow) measurement. The modified spectrum is then inverse Fourier transformed to produce the dynamic image. The term keyhole stems from the idea that peeking through a keyhole of data, the entire k-space can be viewed [Atlas 1996]. The keyhole method can be utilized in applications where a specific region is imaged repeatedly for study and a reduction in imaging time is essential.

Denoting the high resolution reference k-space data by $F_R(k_x, k_y)$, the high resolution k-space data of the dynamic image by $F_D(k_x, k_y)$ and the combined data set generated from applying the keyhole method by $F_C(k_x, k_y)$, the keyhole method can be mathematically formulated in the following manner [Hu 1994]:

$$F_C(k_x, k_y) = F_R(k_x, k_y) + [F_D(k_x, k_y) - F_R(k_x, k_y)]\Pi(\frac{k_x}{W_{k_x}}, \frac{k_y}{W_{k_y}}), \tag{7.1}$$

where $\Pi(\frac{k_x}{W_{k_x}}, \frac{k_y}{W_{k_y}})$ is the 2-D box function with extent W_{k_x} and W_{k_y} in the k_x and k_y direction respectively. Note that the second term of Eqn. 7.1 is simply the low spatial frequency information of the difference between the reference and dynamic image (i.e.

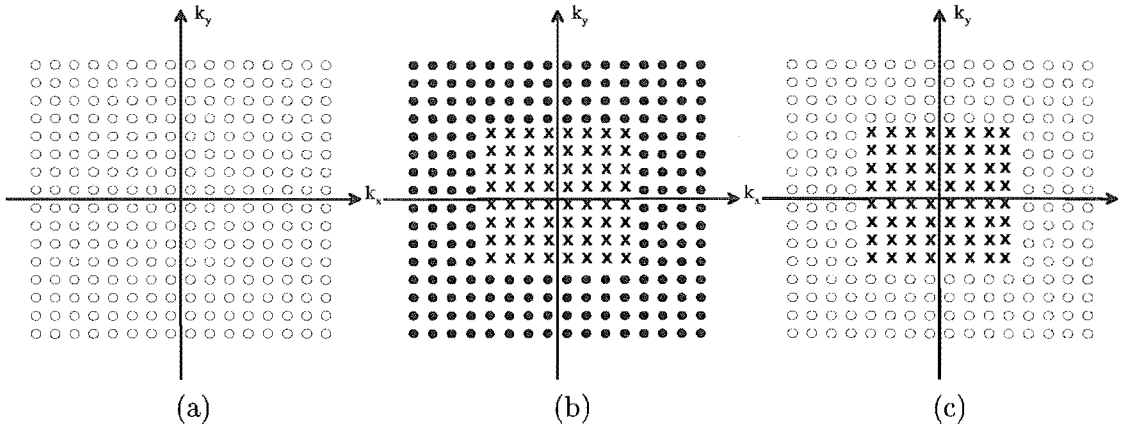


Figure 7.3 Keyhole method: (a) complete data set of reference image; (b) limited data set acquired during dynamic session; (c) combined data set where ∘ represents a collected datum, • represents a missing datum which is set to zero and × represents a datum collected during the dynamic imaging.

the change). Applying the inverse Fourier transform, Eqn. 7.1 can be written as

$$f_C(x,y) = f_R(x,y) + [f_D(x,y) - f_R(x,y)] \star\star W_{k_x} W_{k_y} sinc(W_{k_x}x) sinc(W_{k_y}y), \quad (7.2)$$

where $\star\star$ denotes 2-D convolution, $sinc(x) = \frac{sin(\pi x)}{\pi x}$, $f_R(x,y), f_D(x,y)$ denote the high resolution reference and dynamic image respectively and $f_C(x,y)$ denotes the combined image. This shows that only the difference is degraded by the convolution with the *sinc* function, resulting in the ringing or truncation artifact.

This method was applied to the situation where a follow-up of a growing tumour, representing a region of interest (ROI), in the brain was simulated. Fig. 7.4(a) shows an MR image of a brain with a small tumour (for instance when it was first discovered), which will be referred to as the reference image, or small tumour image, (with an overall SNR of $30dB$), and Fig. 7.4(b) shows the MR image of the same brain but with a larger tumour, referred to as the big tumour image, (again with an overall SNR of $30dB$). During the follow-up session, instead of acquiring the full data set of the image shown in Fig. 7.4(b), only spatial frequencies in the central region of the corresponding k-space were collected (i.e. the lowest 31 spatial frequencies vertically and the lowest 71 spatial frequencies horizontally). The spectrum of the reference image was then modified by substituting the central region of its k-space by the limited data set collected during the follow-up session. The inverse Fourier transform was then applied on the modified spectrum, resulting in the image shown in Fig. 7.4(c). Note that without applying the keyhole method, the image resulting from collecting an incomplete data set at the follow-up session would be the same as the image shown in Fig. 7.2(b). Fig 7.4(d) shows a profile along the tumour (indicated by the red line in Fig. 7.4(c)) of each the high resolution big tumour image, the low resolution big tumour image and the recovered image using the keyhole method. Note that the edges of the original (smaller)

tumour are clearly visible

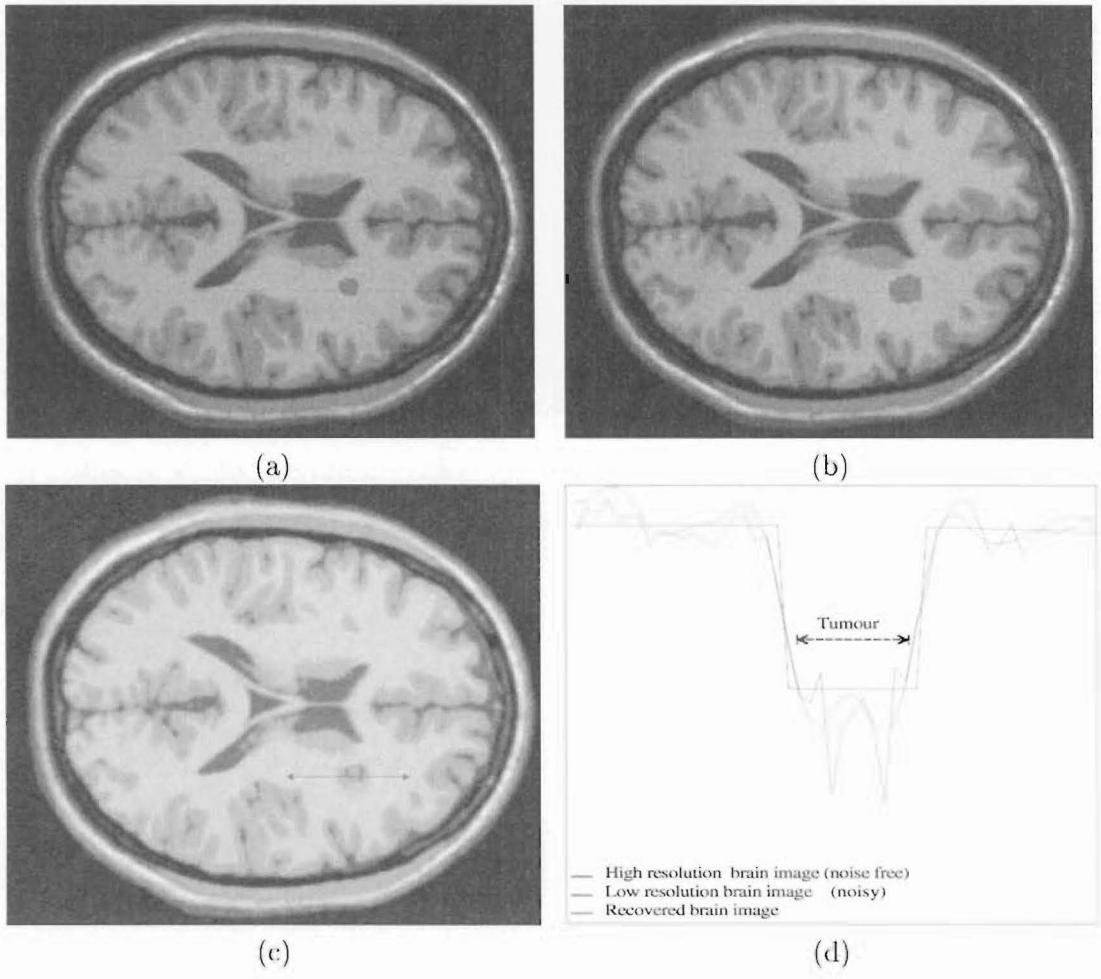


Figure 7.4 Keyhole method: (a) reference image; (b) high resolution image of follow-up session; (c) recovered image obtained from applying the keyhole method; (d) profiles.

A simple illustration of the application of the keyhole method on a 1-D example is shown in Fig. 7.5, where (a) represents the dynamic (e.g. big tumour) image, and (b) represents the reference (small tumour) image. Fig. 7.5 (e) is the resultant image obtained from convolving the sinc function ($W_{k_x} sinc(W_{k_x} x)$) shown in (d) with the difference image shown in (c). The resultant (combined) image from the keyhole method is shown in Fig. 7.5(f), which is simply the sum of (e) and (b) and which shows similarity to the result shown in Fig. 7.4(d).

The recovered image obtained from utilizing the keyhole method has a slight ringing artifact and is much sharper than that obtained from applying the zero filling method. However, it provides little extra information and therefore is not suitable for use in applications where fine details are essential [Jones *et al.* 1993]. The keyhole method is a simple technique that considerably reduces imaging time per dynamic image, is simple to incorporate in the existing MRI equipment, provides three to fivefold gains

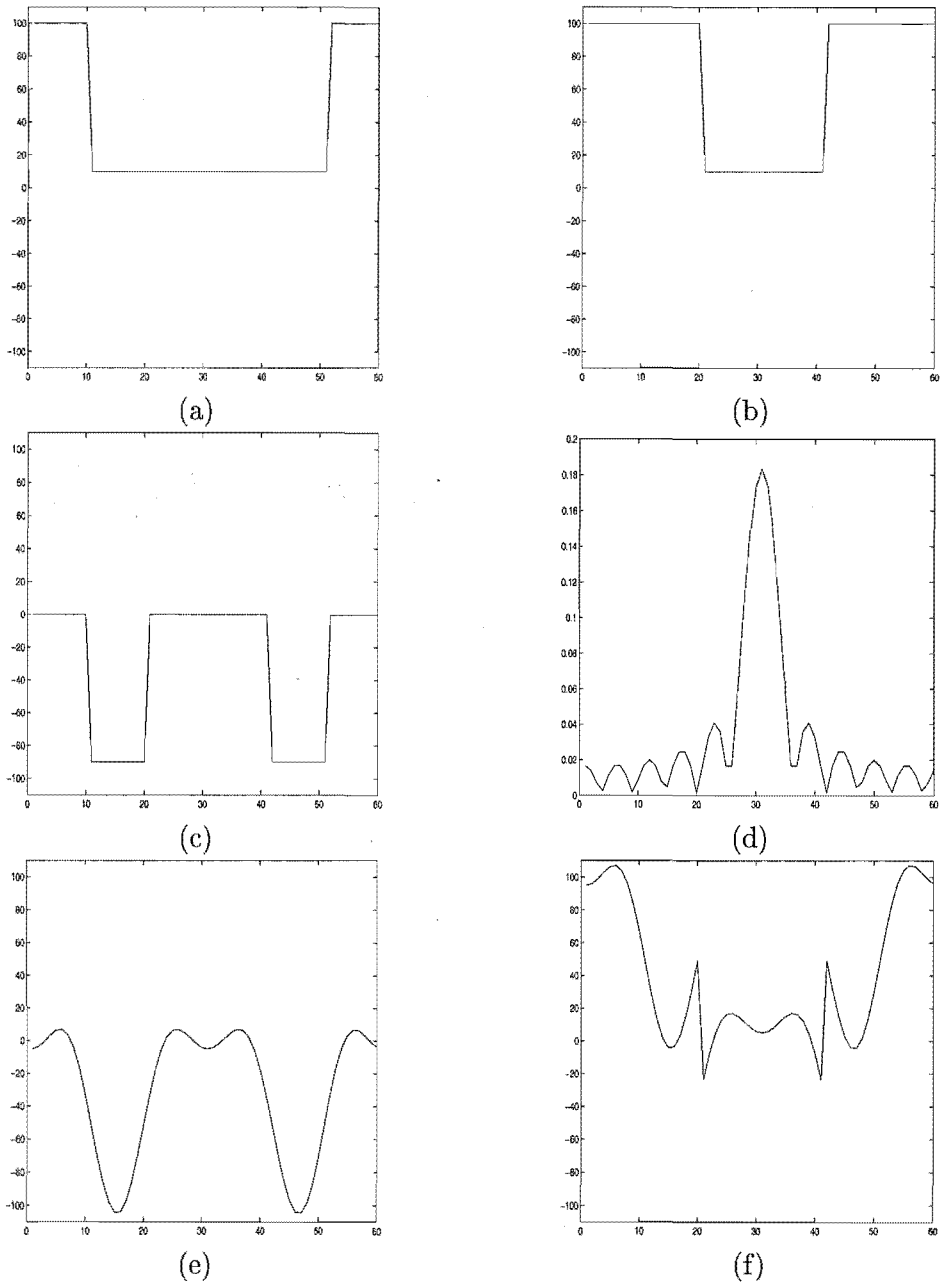


Figure 7.5 Simulation of the keyhole method in 1-D: (a) f_D; (b) f_R; (c) $f_D - f_R$; (d) $|W_{k_x} sinc(W_{k_x} x)|$; (e) $f_D - f_R$ convolved with (d); (f) f_C .

in imaging time, and can be applied in cases where the fine details regarding changes are not so essential.

7.4 REDUCED FIELD OF VIEW METHOD

Another method known as the reduced field of view (rFOV) method, presented by Hu [Hu *et al.* 1994], was suggested to be used for dynamic imaging. Changes localized to a small area of the entire field of view (explained in section 2.5.3) occur in many dynamic applications, such as the insertion of a catheter or a biopsy needle. In this case most of the changes occur in the region containing the catheter or needle. Full sets of data of the entire field of view acquired for each dynamic image will contain a substantial amount of redundant data. Since imaging a reduced field of view reduces the amount of data which needs to be acquired, temporal resolution can be increased. In this method a full data set of a reference image prior to any dynamic change (e.g. before inserting the catheter) is acquired as shown in Fig. 7.6(a). During the dynamic study, a set of data is acquired by sampling densely in the frequency encoding direction and sparsely in the phase encoding direction (since, as explained in section 2.3, the imaging time is proportional to the number of phase encodes). This sparse sampling in the phase encoding direction is achieved by dividing the fully resolved data into r groups, where r is known as the FOV reduction factor. For example if $r = 4$, then the four groups of interleaved phase encoded samples are (0,4,...,N-4), (1,5,...,N-3), (2,6,...,N-2) and (3,7,...,N-1) as shown in Fig. 7.6(b), where N is the number of phase encodes. Note that the amount of undersampling relates to the desired reduction in the FOV which is based on prior knowledge of the size of the area to be imaged. If the inverse Fourier transform was applied on the acquired undersampled image, an aliased image with a reduced FOV would be generated, which is not useful. Instead a difference image with a reduced FOV is then generated by inverse Fourier transforming the difference between the undersampled spectrum and the corresponding k-lines of the reference image. The ROI in the difference image may be shifted from its correct position and needs to be repositioned back to its correct location with respect to the entire field of view. This shift is based on the prior knowledge of the location of ROI with respect to the entire image. The dynamic image is then produced by inverse Fourier transforming the sum of the spectrum of the repositioned difference image and the spectrum of the reference image. This procedure is repeated for each dynamic image. An important point to note is that this type of interleaving phase encoded samples will result in $r - 1$ groups not centred in the k-space, resulting in a linear phase roll in the difference image. Applying a linear phase correction to the dynamic images can be achieved by adding the factor $\frac{2\pi ij}{N}$ to the phase of the dynamic image, where $i = 0, \ldots, \frac{N}{r} - 1$ (pixel number in the reduced FOV) and $j = 0, \ldots, r - 1$ (group number).

In order to mathematically formulate the rFOV method, the high resolution refer-

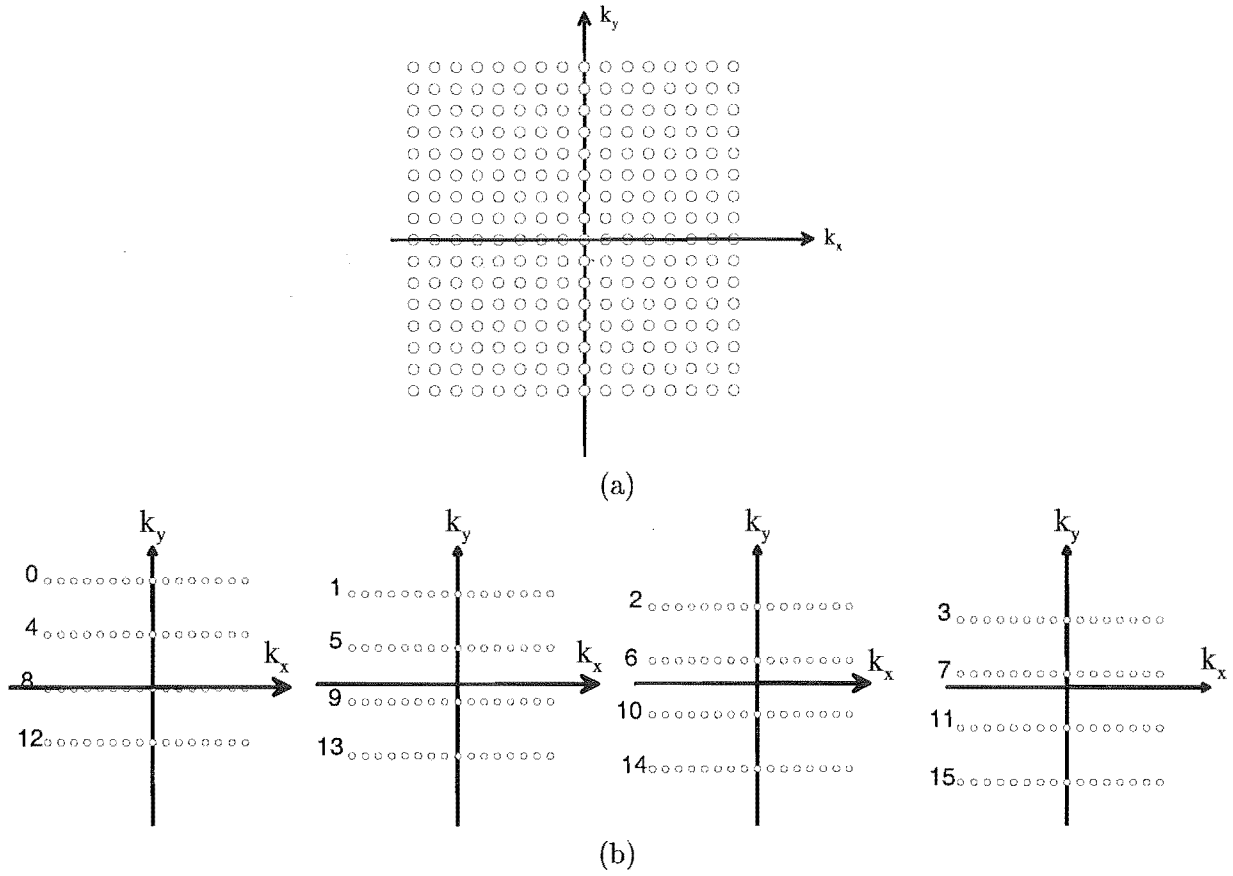


Figure 7.6 (a) A complete k-space data set; (b) rFOV acquired data sets for dynamic imaging.

ence image prior to any dynamic change is denoted by f_R. The high resolution dynamic image is denoted by f_D and the dynamic difference or change between f_R and f_D is denoted by Δf. Therefore one can write

$$\Delta f = f_D - f_R. \tag{7.3}$$

Since the following formulation is with respect to Δf, the Δf will be reduced to f for notational simplicity. For the case $r = 2$, the two groups of interleaved samples are $(0, 2, \ldots, N-2)$, denoted by $\mathcal{G}^e$, and $(1, 3, \ldots, N-1)$, denoted by $\mathcal{G}^o$. For the $\mathcal{G}^e$ case, an undersampled spectrum $\hat{F}^e$ can be generated by using the undersampling pattern S^e such that

$$\hat{F}^e_{p,q} = S^e_{p,q} F_{p,q} = \frac{1}{2}(1 + (-1)^p) F_{p,q}, \quad p, q = 0, \ldots, N-1. \tag{7.4}$$

Following the derivations in chapter 6, we have

$$\begin{aligned}
\hat{f}^e_{n,m} &= \frac{1}{2}(f_{n,m} + f_{n+\frac{N}{2},m}) \quad n, m = 0, \ldots, N-1 \qquad (7.5)\\
&= \frac{1}{N^2} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} \hat{F}_{p,q} w_N^{np} w_N^{mq} \\
&= \frac{1}{N^2} \sum_{p=0,2}^{N-1} \sum_{q=0}^{N-1} \hat{F}_{p,q} w_N^{np} w_N^{mq} \\
&= \frac{1}{N^2} \sum_{p'=0}^{\frac{N}{2}-1} \sum_{q=0}^{N-1} \hat{F}_{2p',q} w_N^{2np'} w_N^{mq} \\
&= \frac{1}{N^2} \sum_{p'=0}^{\frac{N}{2}-1} \sum_{q=0}^{N-1} \hat{F}_{2p',q} w_{\frac{N}{2}}^{np'} w_N^{mq}. \qquad (7.6)
\end{aligned}$$

Alternatively, the spectrum $\tilde{F}^e$ can be produced by $\mathcal{G}^e$ by simply arranging the even rows of $\hat{F}^e$ in a new image $\frac{N}{2} \times N$, illustrated in Fig. 7.7. The corresponding image $\tilde{f}^e$ can be written as

$$\begin{aligned}
\tilde{f}^e_{n',m} &= \frac{1}{\frac{N}{2}} \frac{1}{N} \sum_{p'=0}^{\frac{N}{2}-1} \sum_{q=0}^{N-1} \tilde{F}^e_{p',q} w_{\frac{N}{2}}^{n'p'} w_N^{mq} \\
&= \frac{1}{\frac{N}{2}} \frac{1}{N} \sum_{p'=0}^{\frac{N}{2}-1} \sum_{q=0}^{N-1} \hat{F}^e_{2p',q} w_{\frac{N}{2}}^{n'p'} w_N^{mq} \quad n' = 0, \ldots, \frac{N}{2} - 1. \qquad (7.7)
\end{aligned}$$

Comparing Eqns. 7.6 and 7.7, one can write

$$\tilde{f}^e_{n',m} = 2\hat{f}^e_{n,m} \quad n' = n = 0, \ldots, \frac{N}{2} - 1. \tag{7.8}$$

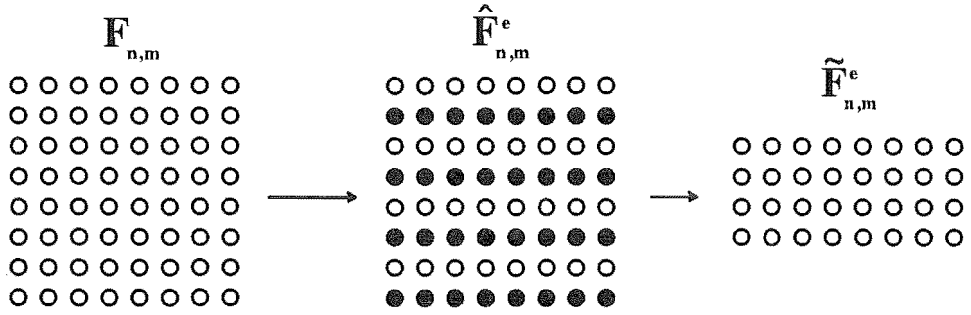


Figure 7.7 Illustration of the acquiring of dynamic data in the rFOV method where ○ represents a collected datum and • represents a missing datum.

Due to the periodicity of both $\tilde{f}^e$ and $\hat{f}^e$ with respect to n over $\frac{N}{2}$ we have

$$\begin{aligned} \tilde{f}^e_{n,m} &= 2\hat{f}^e_{n,m} \quad \forall n,m \\ &= f_{n,m} + f_{n+\frac{N}{2},m}. \end{aligned} \tag{7.9}$$

Fig. 7.8 shows three different cases of the aliased image that could be produced when the extent of the ROI in the difference image $f_{n,m}$ is less than $\frac{N}{2}$. In each of the cases illustrated $\tilde{f}^e_{n,m}$ is inserted into the top half of the reconstruction. For the first case, when $f_{n,m}$ is zero over half its period (i.e. $f_{n,m} = 0$, for $n = \frac{N}{2}, \ldots, N-1$), the reconstructed image $\tilde{f}^e_{n,m} = f_{n,m}$ as shown in Fig. 7.8(a). If $f_{n,m} = 0$ for $n = 0, \ldots, \frac{N}{2} - 1$, then $\tilde{f}^e_{n,m} = f_{n+\frac{N}{2},m}$ as shown in Fig. 7.8(b) and needs to be shifted back to its correct location. Fig. 7.8(c) portrays the case when the ROI extends into both halves, in which case the aliasing causes the ROI to be split in $\tilde{f}^e_{n,m}$ (A and C). Knowledge of the ROI's location is therefore necessary to shift part of the image (C and D in Fig. 7.8(c)). Note that if the extent of the ROI is greater than $\frac{N}{2}$, then in Fig. 7.8(c), region A and C in $\tilde{f}^e_{n,m}$ will alias over each other making it impossible to recover the ROI. This can be extended to cases with $r > 2$, where the extent of ROI should be less than N/r.

Considering the other part of the spectrum, $\mathcal{G}^o$, we have

$$\hat{F}^o_{p,q} = S^o_{p,q} F_{p,q} = \frac{1}{2}(1 - (-1)^p) F_{p,q}, \quad p,q = 0, \ldots, N-1, \tag{7.10}$$

and the corresponding image can be written as

$$\hat{f}^o_{n,m} = \frac{1}{2}(f_{n,m} - f_{n+\frac{N}{2},m}) \quad n,m = 0, \ldots, N-1 \tag{7.11}$$

$$= \frac{1}{N^2} w_N^n \sum_{p'=0}^{\frac{N}{2}-1} \sum_{q=0}^{N-1} \hat{F}_{2p'+1,q} w_{\frac{N}{2}}^{np'} w_N^{mq}. \tag{7.12}$$

The spectrum $\tilde{F}^o$, produced by $\mathcal{G}^o$ is simply the odd rows of $\hat{F}^o$ arranged in a new

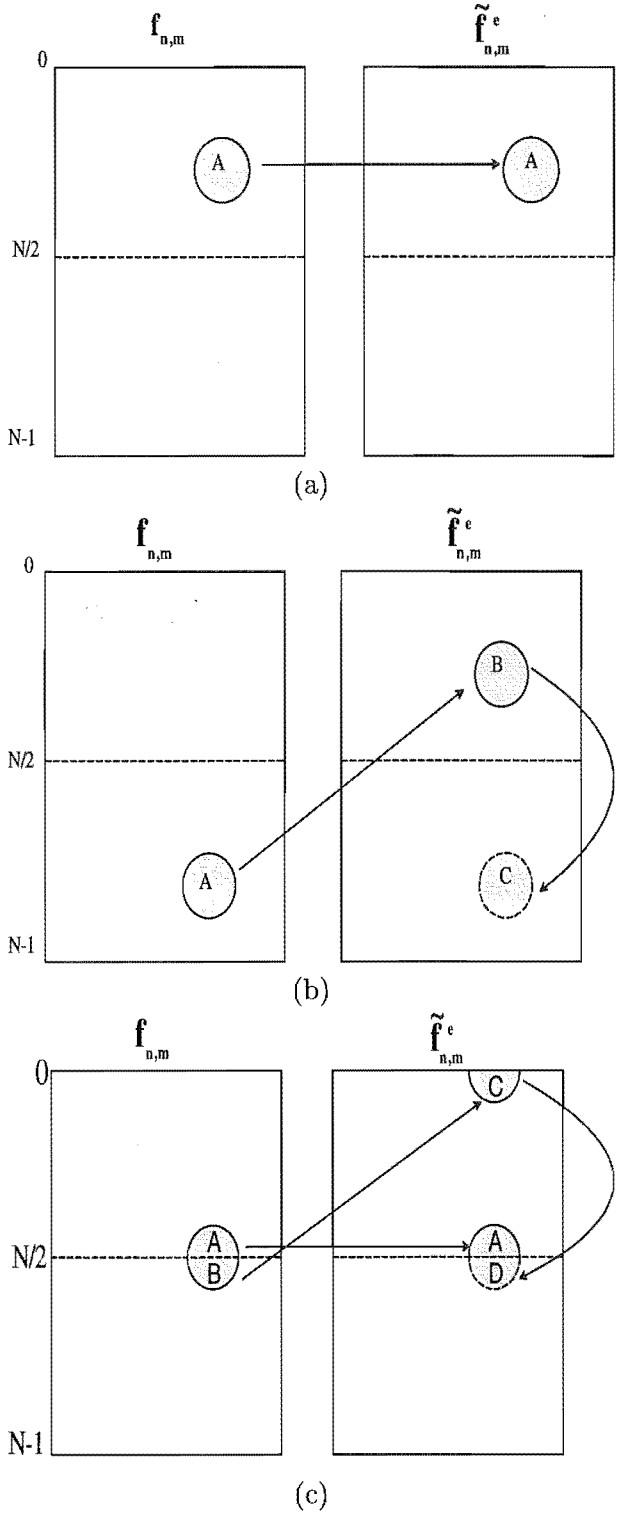


Figure 7.8 Illustration of rFOV with $r = 2$: $\tilde{f}^e_{n,m}$ if (a) $f_{n,m} = 0$, for $n = \frac{N}{2}, \ldots, N-1$; (b) $f_{n,m} = 0$ for $n = 0, \ldots, \frac{N}{2} - 1$; (c) $f_{n,m} \neq 0$ for $n = 0, \ldots, N-1$.

image $\frac{N}{2} \times N$ and the corresponding image $\tilde{f}^o$ can be written as

$$\tilde{f}^o_{n',m} = \frac{1}{\frac{N}{2}} \frac{1}{N} \sum_{p'=0}^{\frac{N}{2}-1} \sum_{q=0}^{N-1} \hat{F}^o_{2p'+1,q} w_{\frac{N}{2}}^{n'p'} w_N^{mq} \quad n' = 0, \ldots, \frac{N}{2} - 1. \tag{7.13}$$

Comparing Eqns. 7.12 and 7.13, we have

$$\begin{aligned} \tilde{f}^o_{n,m} &= 2 w_N^{-n} \hat{f}^o_{n,m} \quad \forall n, m \\ &= w_N^{-n} (f_{n,m} - f_{n+\frac{N}{2},m}). \end{aligned} \tag{7.14}$$

Similar derivations can be made for different values of r. Eqns. 7.9 and 7.14 indicate that any change will either remain in its position or be aliased into a different position depending on the original location of the difference. Therefore in order to reposition the difference, the original location of the difference needs to be known. Note the number of terms in Eqns. 7.9 and 7.14 (i.e. the aliased portions) increases with r.

The reduced field of view method was applied on the MR brain with tumour image. The high resolution noisy small tumour image, shown in Fig. 7.4(a), was used as the reference image. The undersampled reduced field of view data was obtained by undersampling the high resolution noisy large tumour image, shown in Fig. 7.4(b), by missing out every other k-space line data along the phase encoding direction. The inverse Fourier transform of the acquired data is shown in Fig. 7.9(a), which illustrates the aliasing effect due to the undersampling in the phase encoding direction. The difference image was then generated by inverse Fourier transforming the difference between the acquired reduced field of view data and the spectrum of the reference image and is shown in Fig. 7.9(b). The dynamic image, shown in Fig. 7.9(c), is then produced by correctly repositioning the reduced field of view with respect to the entire field of view and then inverse Fourier transforming the sum of the spectra of both the re-positioned data and the reference image. Fig. 7.9(d) shows a profile along the tumour (indicated by the red line in Fig. 7.9(c)) of each of the high resolution brain with tumour image and the recovered image using the reduced field of view method.

In contrast to the results obtained with the keyhole method, the reduced field of view method does not degrade the spatial resolution of the dynamic information, and therefore can be applied in situations where high resolution dynamic images are required. However one disadvantage of this method is that it requires prior information regarding both the extent (to chose r) and position of the ROI (to allow for correct repositioning of the ROI). In interventional procedures, the position of the reduced field of view can be estimated, however the quality of the images will depend on the error of the operator's estimate [Kyriakos *et al.* 1997]. Another major disadvantage is the assumption that the changes in the dynamic image are physically restricted within an

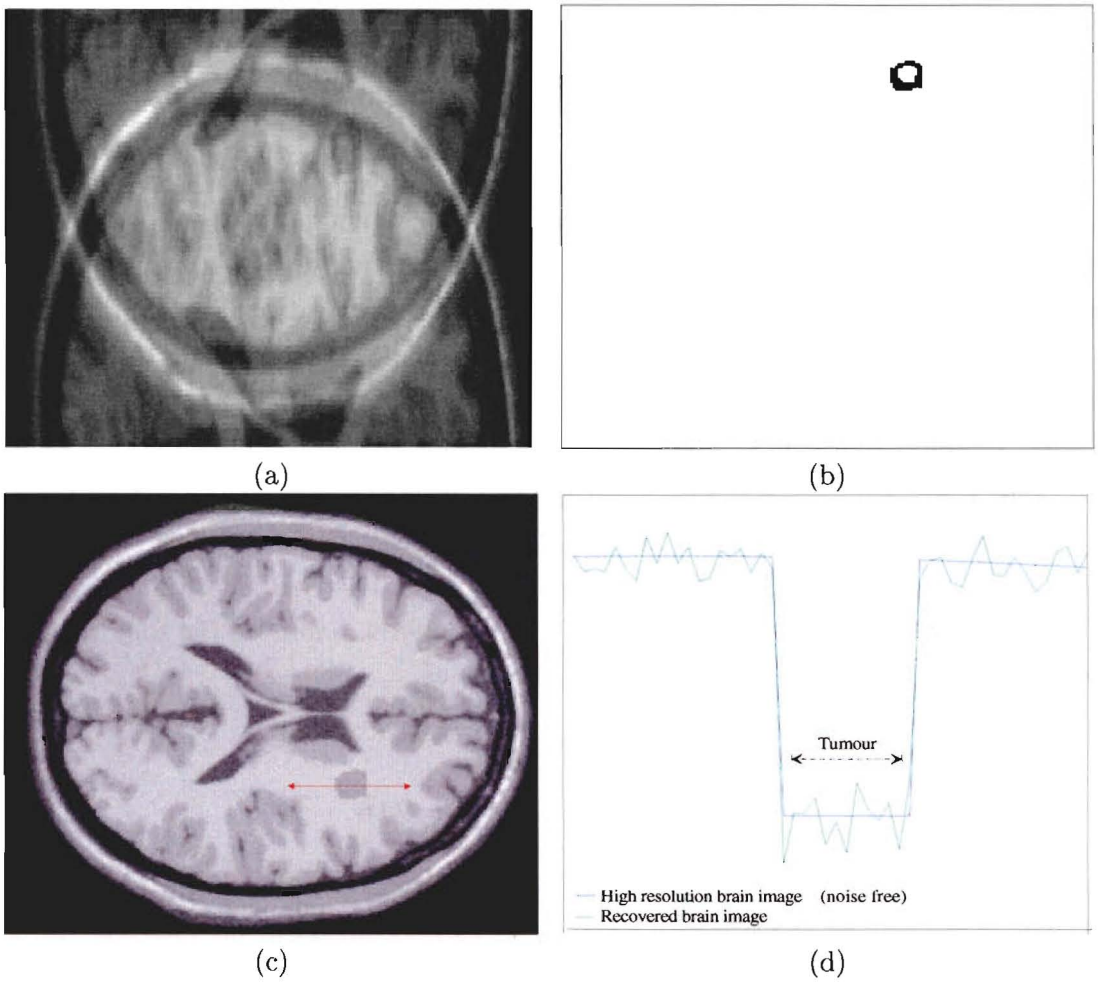


Figure 7.9 Reduced field of view method (a) undersampled image; (b) difference image; (c) resultant dynamic image; (d) profiles.

a priori known reduced field of view, which results in aliasing inside the reduced field of view of any changes that occur outside the reduced field of view [Hu and Parrish 1994]. Therefore the reconstructed dynamic image resulting from the use of this technique contains significant artifacts within the reduced field of view. This limitation makes this method inapplicable to some specific applications such as: abdominal and cardiac imaging (due to the respiratory motion), and brain imaging when there is slight bulk head motion.

7.5 ITERATIVE RECOVERY OF ROI

A new method, known as the iterative recovery of ROI (iROI), is presented here, which overcomes some of the drawbacks of the above mentioned methods. This method provides sharper images than those obtained with the reduced acquisition or keyhole methods and does not suffer from the drawback of having artifacts within the ROI due to changes outside the ROI as in the case of the rFOV method, though it does still require an approximate prior knowledge of the extent and location of ROI.

This method is used in a situation where a fully resolved image f_R, referred to as the reference image, is available, and a follow-up image of the ROI $\bar{f}$ is required by obtaining a low resolution image g of the entire FOV. The G-P algorithm is used to superresolve the region of interest by utilizing the information from the fully resolved image f_R. The imaging process can be formulated as

$$\bar{A}\bar{f} + \hat{A}f_R = G \tag{7.15}$$

$$\Rightarrow \bar{A}\bar{f} = G - \hat{A}f_R = \bar{G}, \tag{7.16}$$

where $\hat{A} = \xi\mathcal{F}(1-\bar{\gamma})$, $\bar{A} = \xi\mathcal{F}\bar{\gamma}$, ξ is the frequency domain constraint, $\bar{\gamma}$ is the space domain constraint for the region of interest, G is the spectrum of the image g, and $\mathcal{F}$ is the forward Fourier transform [Alwesh *et al.* 1999]. The iterative steepest descent method, which is equivalent to the G-P algorithm for steplength $\alpha^k = \frac{1}{2}$ [Connolly *et al.* 1995], can then be used to solve the above linear system which leads to the following formulation (following the procedure of section 3.5)

$$\bar{f}^{k+1} = \bar{f}^k + \bar{\gamma}\mathcal{F}^{-1}\xi\bar{G} - \bar{\gamma}\mathcal{F}^{-1}\xi F\bar{f}^k. \tag{7.17}$$

A study of the application of this method on two types of low resolution images was made. For the first type, a fully resolved small tumour image, shown in Fig. 7.4(a), was used as the reference image (i.e. f_R) and a follow-up of the ROI (i.e. the tumour) was needed. A low resolution image was then constructed (shown in Fig 7.2), by truncating the fully resolved spectrum of the large tumour image, shown in Fig. 7.4(b) where its spectrum was truncated to the lowest 31 spatial frequencies vertically and the lowest 71

spatial frequencies horizontally (i.e. G). The iterative ROI method was then applied to the bandlimited image with $\bar{\gamma}$ set to the exact ROI (tumour) mask, referred to as mask 1, shown in Fig. 7.10(a). Fig. 7.10(b) shows the recovered image obtained by applying the iterative ROI method within the ROI, then inserting this within the original high resolution image. It shows an impressive recovery of the low resolution image. Fig. 7.10(c) shows a profile across the tumour (indicated by the vertical red line in Fig. 7.10(b), of the high resolution large tumour image, of the corresponding low resolution image, and of the recovered image, which clearly shows the improvement in the recovery of the edges of the tumour.

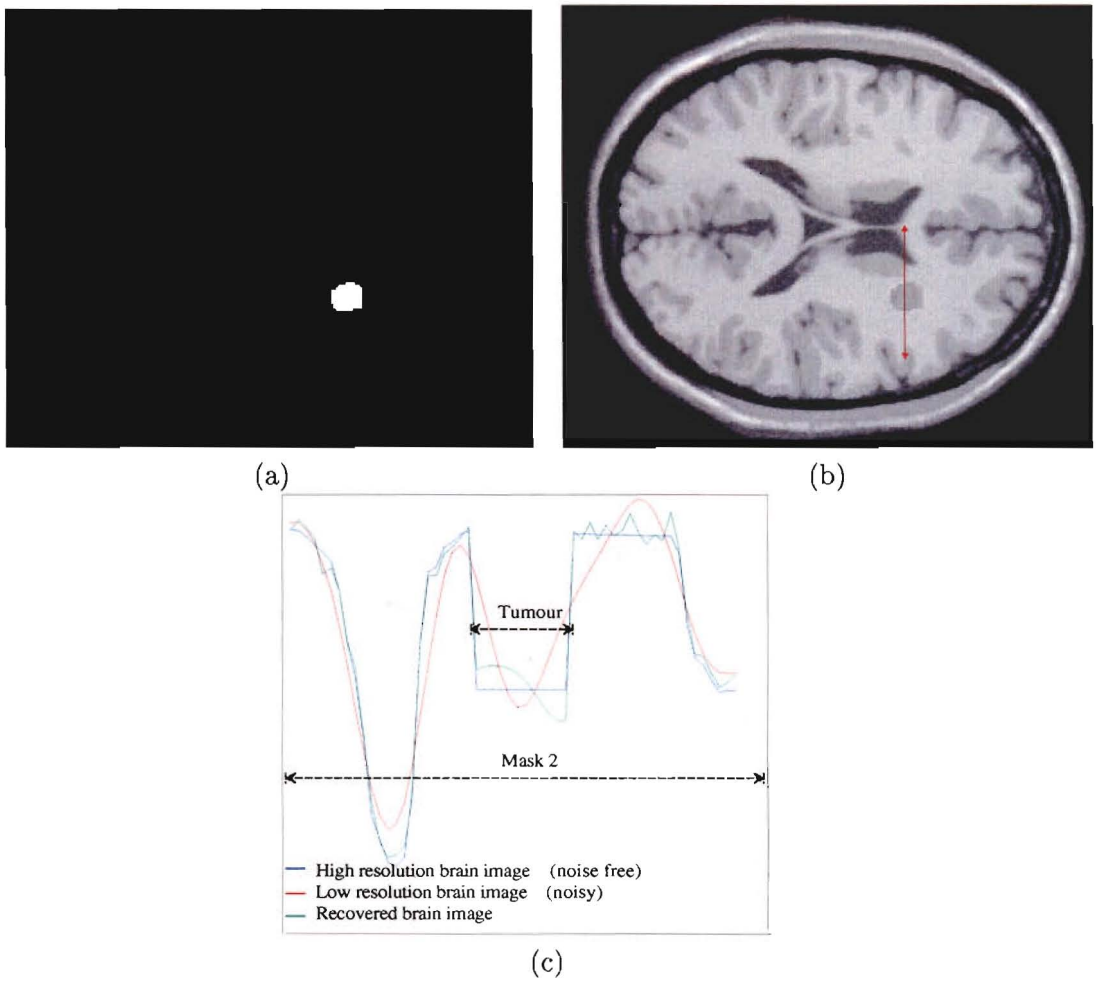


Figure 7.10 iROI method: (a) mask 1; (b) recovered image; (c) profiles.

In practice, the exact region of interest is never known but a reasonable approximation to it can be deduced from the high resolution reference image. The iterative ROI method was again applied on the bandlimited brain with the large tumour image, with $\bar{\gamma}$ set to the mask shown in Fig. 7.11(a), referred to as mask 2. The recovered image in this case, shown in Fig. 7.11(b), is quite blurred within the ROI. To recover the edges (especially of the tumour), the total variation method, discussed in section

5.6.1, with $\lambda = 20$ and $\alpha = 0.005$ was applied on the recovered image. This particular constraint is useful here due to the relatively low total variation of the tumour. The resultant image, shown in Fig. 7.11(c), shows a significant improvement and the edges are quite sharp. Note the ringing artifact around the edges of the mask, which is due to the bandlimiting operation which corresponds to convolving $(1 - \bar{\gamma})f$ with a sinc function when computing $\bar{G}$. However, the ringing artifact does not seem to compromise information regarding the tumour. Comparing the results of the various methods, the iterative ROI method provides a sharper, more resolved image compared to the results obtained from both the reduced acquisition and keyhole methods. However the recovered image obtained from the reduced field of view method is better (it requires no total variation method and does not suffer from the ringing artifact of the edges) than that obtained from the iterative ROI, but has the major limitation that changes have to be within the rFOV, and if this condition is violated artifacts will result.

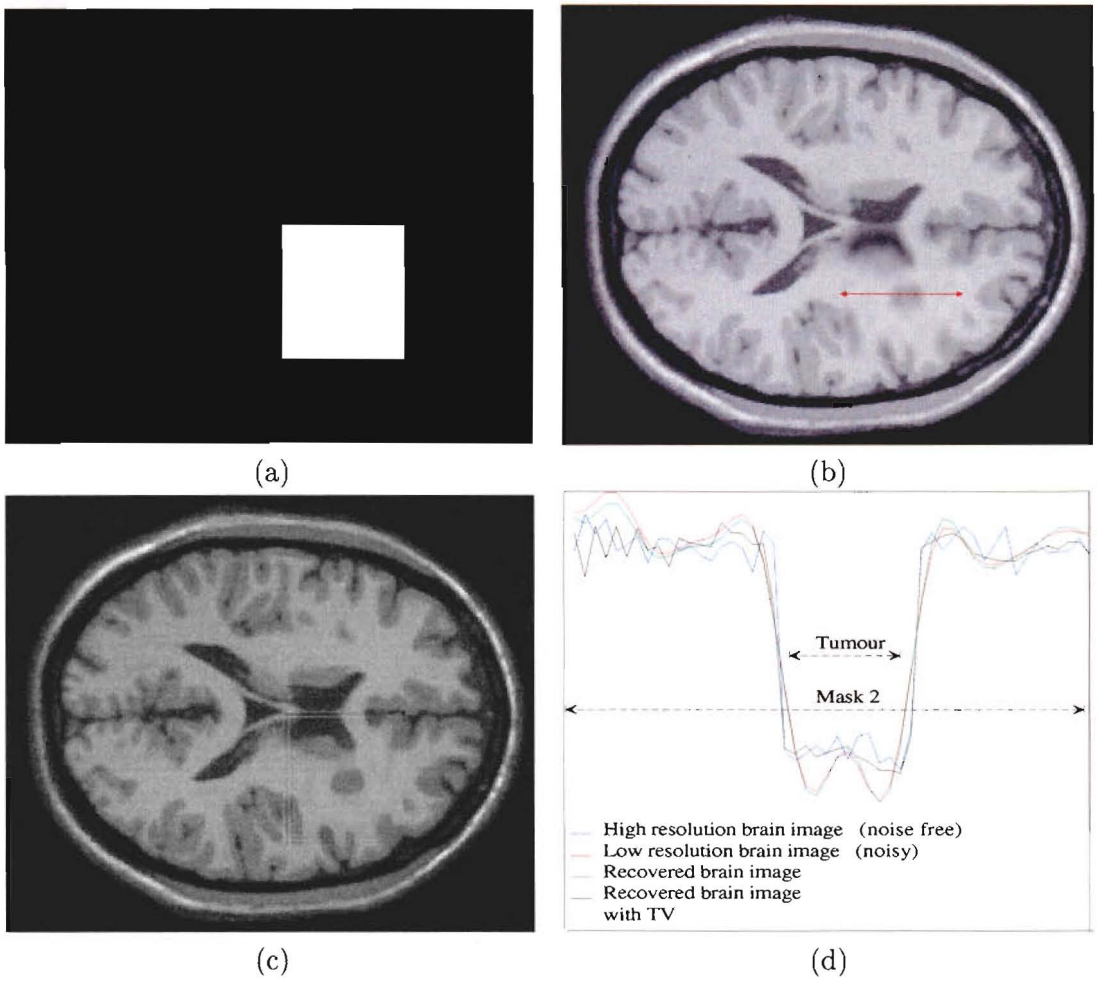
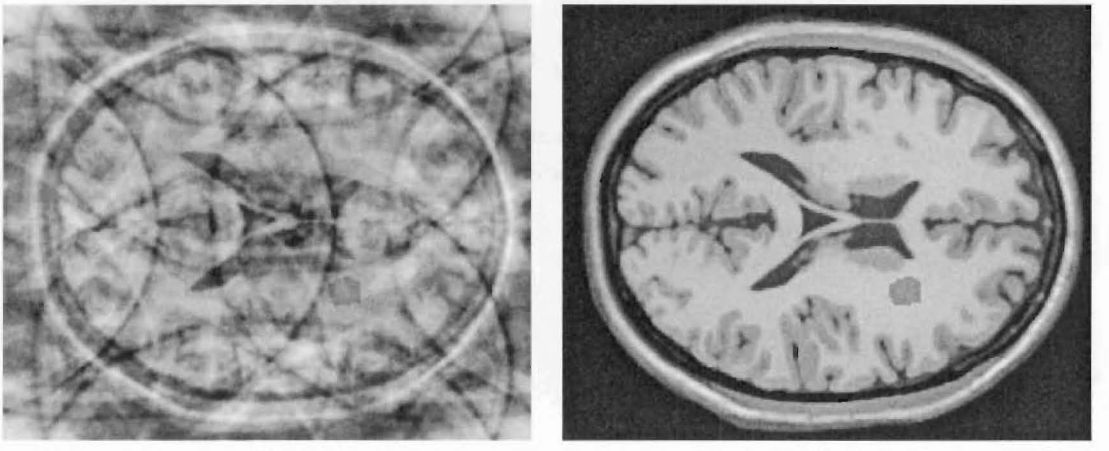


Figure 7.11 iROI method: (a) mask 2; (b) recovered image; (c) recovered image from total variation method with $\lambda = 20$ and $\alpha = 0.005$; (d) profiles.

The second type of low resolution images were constructed by undersampling the

spectrum of the large tumour image by multiplying it with an undersampling pattern, $S^6_{p,q}$ (see chapter 6), resulting in the aliased image shown in Fig. 7.12(a). The recovered image shown in Fig. 7.12(b) was then computed by applying Eqn. 7.17 on the low resolution aliased image with the fully resolved small tumour image as a reference and $\bar{\gamma}$ set to mask 2. The aliasing is completely removed from the ROI, and the edges of the tumour are very clearly defined. Similar results were obtained for the different undersampling patterns discussed in chapter 6.



(a) (b)

Figure 7.12 (a) Undersampled large tumour image; (b) recovered image using mask 2.

Another form of the above method, where the reference image utilized is a low resolution image, was studied. In this case, the high resolution dynamic image (i.e. with tumour) is considered to be the sum of the high resolution reference image f_R (with no tumour) and a difference image $\bar{f}$. Therefore the collected bandlimited data set (G) of the dynamic image can be written as

$$\begin{aligned} A(f_R + \bar{f}) &= G \Rightarrow \\ A\bar{f} &= G - Af_R. \end{aligned} \tag{7.18}$$

where $A = \xi\mathcal{F}\gamma$. Therefore Af_R is simply the bandlimited measured spectrum of the reference image (with no tumour). Since $\bar{f}$ is nonzero only within $\bar{\gamma}$, $A\bar{f} = \bar{A}\bar{f}$ where $\bar{A} = \xi\mathcal{F}\bar{\gamma}$. Eqn. 7.18 can therefore be written as

$$\bar{A}\bar{f} = G - Af_R = \bar{G}. \tag{7.19}$$

which can be solved for $\bar{f}$ by using Eqn. 7.17, with $\bar{G}$ as defined in Equ. 7.19. The above method was applied on the bandlimited large tumour image, with the reference image being the bandlimited no tumour image. In order to enhance the edges, the total variation method was applied with $\lambda = 30$. The recovered image is shown in Fig. 7.13, which shows considerable recovery of the tumour compared to the bandlimited image

shown in Fig. 7.2(b). Note also the absence of ringing artifacts in this case.

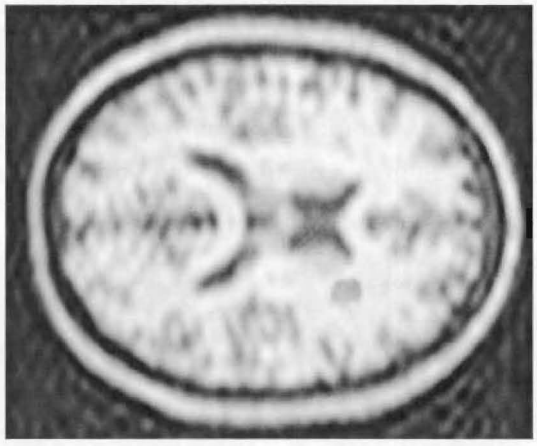


Figure 7.13 Recovered image using mask 2, a low resolution no tumour image as a reference image and total variation method with $\lambda = 30$.

7.6 CONCLUSION

In this chapter several methods which reduce imaging time have been discussed. These methods were discussed in the context of applying them to dynamic imaging or for follow-up study purposes. Both the reduced acquisition and keyhole methods are very simple methods that can be easily incorporated in existing MRI equipment and which reduce scanning time significantly. The reduced acquisition method requires no high resolution reference image unlike the keyhole method. They are both not useful in applications where fine details are important. The rFOV method provides extremely good results if the changes are restricted within a specific reduced FOV. This limitation restricts the application of the method, for example excluding cardiac or abdominal imaging.

The iROI method was presented in two forms. In the first form, a high resolution reference data is required, and if the mask utilized is not the exact size of the ROI then the total variation method needs to be applied to enhance the edges. The method (in unmodified form) also suffers from a ringing artifact that is noticeable close to the edges of the mask. The second form of the iROI requires only a low resolution reference data and also provides good reconstruction of the ROI. Neither form of the iROI suffers from the limitation of the rFOV method and the second form of the iROI does not require a high resolution reference image. It is important to mention that the ISNR curves obtained from applying the iROI algorithm in both forms were not levelled off except for when applied to an undersampled image and when utilizing the total variation method. Therefore the corresponding recovered images were obtained at iterations with the maximum ISNR value.

Chapter 8

PARAMETER ESTIMATION METHODS AND THE NONCONVEX LEVEL PENALTY FUNCTION

8.1 INTRODUCTION

Magnetic resonance images, as discussed in chapter 2, are not measured directly but instead are obtained by utilizing the Fourier imaging technique, where spatially resolved information regarding the imaged object is frequency and phase encoded in the acquired data. The image is then reconstructed by applying the 2-D inverse FFT (equivalent to applying the 1-D inverse FFT on the rows followed by the 1-D inverse FFT on each column independently). In theory, if enough data is collected and the data is not corrupted by error, then the images can be reconstructed perfectly. However in practice, due to time and physical limitations, a limited amount of data is collected, resulting in a loss of resolution and truncation artifacts. Each column and row of the collected data can be considered as an infinite row or column multiplied by a rectangular window which corresponds to the convolution of the DFT of the infinite row or column by a *sinc* function resulting in Gibb's oscillation [Haacke *et al.* 1989a]. A reduction in the sidelobes of the convolving function can be achieved by choosing a smoother window; this results in a reduction of the Gibb's oscillation but also an undesirable reduction in resolution.

In recent years, parametric estimation methods (i.e. non-Fourier based) have been proposed, where a chosen model (depending on prior information) can represent a class of image functions with a finite number of unknowns (the parameters). The parameters (rather than the image pixel values) need to be estimated from the measured data. The main factors determining which model to choose are [Liang and Lauterbur 1999]:

- the ability of the model to represent the class of image functions,
- the number of parameters required to model the images; for example two models might yield similar results, however they might differ in the number of model parameters required,
- the stability of the model under small perturbations of the data, and

- the speed and feasibility of the computation of the parameters.

In this chapter, modelling through the use of boxcar functions suggested by Haacke [Liang *et al.* 1989] is discussed. Before discussing the boxcar estimation method, both the singular value decomposition based least squares method and Prony's method need to be described since they will be utilized in the boxcar estimation method.

Another method we introduce that utilizes prior information is the nonconvex level penalty function, where information regarding the image's levels are incorporated within the G-P algorithm. This constraint aims at restricting the pixel values to a set of predetermined levels. Details regarding this constraint and results are discussed in section 8.6.

8.2 SINGULAR VALUE DECOMPOSITION (SVD) BASED LEAST SQUARES METHOD

The SVD method is used to factorize an arbitrary $n \times m$ matrix A with rank r, into the product of orthonormal matrices $n \times n, U = [u_1, u_2, \ldots, u_n]$, $m \times m, V = [v_1, v_2, \ldots, v_m]$ and diagonal matrix $n \times m, \Sigma$ such that

$$A = U\Sigma V^{\mathcal{T}^*} = \sum_{i}^{r} \sigma_i u_i v_i^{\mathcal{T}^*}$$

$$\text{where } \Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \sigma_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & & \ddots & \\ & & & \sigma_r \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right], \tag{8.1}$$

the σ_i are known as the singular values of A (nonnegative square roots of the eigenvalues of $A^{\mathcal{T}^*}A$ and $AA^{\mathcal{T}^*}$, the columns of U are the orthonormal eigenvectors of $AA^{\mathcal{T}^*}$, the columns of V are the orthonormal eigenvectors of $A^{\mathcal{T}^*}A$, and $\mathcal{T}^*$ is the conjugate transpose operator. Many problems require the solution of a set of simultaneous linear equations expressed in the form

$$Ax = b, \tag{8.2}$$

where A is a $n \times m$ matrix, b is a $n \times 1$ vector and x is a $m \times 1$ vector of unknowns. For example in imaging problems, the problem is to find the object x from the collected data b. From Eqn. 8.2, to solve for x the most direct method is $x = A^{-1}b$ (for $n = m$), where A^{-1} is the inverse of matrix A. However if $n > m$, the system is overdetermined and a least squares approach needs to be used. In this case the aim is to find a solution

$\hat{x}$ that minimizes both $\|b - A\hat{x}\|^2$ and $\|\hat{x}\|^2$. This is achieved by computing the Moore-Penrose generalized inverse $A^{\#}$ where [Watkins 1991]

$$A^{\#} = V\Sigma^{\#}U^{T^*}$$

$$\text{where } \Sigma^{\#} = \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & & \ddots & \\ & & \frac{1}{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]. \tag{8.3}$$

The solution obtained with the generalized inverse, $\hat{x} = A^{\#}b$, is equivalent to the direct least squares solution $\hat{x} = (A^{T^*}A)^{-1}A^{T^*}b$, if A is of full rank (i.e. $r = m$). To understand the advantage of using the SVD based least squares method, in practice the collected data is corrupted with noise b_e and the collected data can be written as $\hat{b} = b + b_e$. This results in an error in the solution $\hat{x}$, such that $\hat{x} = x + x_e$. A bound for x_e due to the presence of b_e can be written as [Walsh and Nielsen Delaney 1994]

$$\frac{\|x_e\|}{\|x\|} \leq cond(A)\frac{\|b_e\|}{\|b\|}, \tag{8.4}$$

where $cond(A)$ is the ratio of the largest singular value of A to the smallest singular value of A. Since the SVD method is linear, the error x_e can be written as

$$x_e = V\Sigma^{\#}[U^{T^*}b_e]. \tag{8.5}$$

It is obvious that if some singular values are very small, their contribution to the error x_e is large. Therefore, by setting the very small singular values to zero, the error due to noise can be reduced significantly [Walsh and Nielsen Delaney 1994].

8.3 PRONY'S METHOD

Before discussing the boxcar estimation method, Prony's method needs to be described since it will be utilized in the next section. The classical method of Count de Prony models a sequence of $2P$ observations made at equal intervals by a linear combination of exponential functions. Prony's ingenious method converted the problem to a system of linear equations. If N complex data samples, $x(1), \ldots, x(N)$, were available, then $x(n)$ can be estimated with P exponential terms such that [S. Lawrence Marple 1987]

$$x(n) = \sum_{m=1}^{P} A_m e^{(a_m + i\omega_m)nT + i\theta_m}, \tag{8.6}$$

for $n = 1, \ldots, N$, T is the sampling interval, A_m is the amplitude of the m^{th} exponential, a_m is the damping factor, θ_m is the initial phase and ω_m is the angular frequency. Let $h_m = A_m e^{i\theta_m}$ and $z_m = e^{(a_m + i\omega_m)T}$ and substituting them into Eqn. 8.6 we have

$$x(n) = \sum_{m=1}^{P} h_m z_m^n. \tag{8.7}$$

Since there are $2P$ unknowns $(h_1, \ldots, h_P, z_1, \ldots, z_P)$, the minimum value of N is $2P$. An important key to solving for the parameters is to note that Eqn. 8.7 is the solution to a constant coefficient linear difference equation defined as follows:

$$\Phi(z) = \prod_{m=1}^{P} (z - z_m) = \sum_{m=0}^{P} \rho(m) z^m, \tag{8.8}$$

where $\rho(0) = 1$. Multiplying a shifted version of $x(n)$ (from n to $n+m$) of Eqn. 8.7 by $\rho(m)$ and summing over all $P+1$ products we have

$$\begin{aligned} \sum_{m=0}^{P} \rho(m) x(n+m) &= \sum_{i=1}^{P} h_i \sum_{m=0}^{P} \rho(m) z_i^{n+m} \quad \text{for } 1 \le n \le N - P \\ &= \sum_{i=1}^{P} h_i z_i^n \sum_{m=0}^{P} \rho(m) z_i^m = 0 \end{aligned} \tag{8.9}$$

$$\text{Thus} \quad -x(n) = \sum_{m=1}^{P} \rho(m) x(n+m). \tag{8.10}$$

The equating of the right hand side of Eqn. 8.10 to zero results from the fact that $\Phi(z_i) = \sum_{m=0}^{P} \rho(m) z_i^m$ is evaluated at each of its roots. Eqn. 8.10 can now be written as a set of linear prediction equations such that

$$\begin{Bmatrix} x(2) & x(3) & \ldots & x(P+1) \\ x(3) & x(4) & \ldots & x(P+2) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-P+1) & x(N-P+2) & \ldots & x(N) \end{Bmatrix} \begin{Bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(P) \end{Bmatrix} = - \begin{Bmatrix} x(1) \\ x(2) \\ \vdots \\ x(P) \end{Bmatrix} \tag{8.11}$$

$\rho(m)$ can then be found by using the SVD based least squares procedure presented in section 8.2. Once the ρ's are calculated, the roots z_m in Eqn. 8.8 can be found by using polynomial factoring methods, which in turn are used to find a_i and ω_i according to

$$\begin{aligned} a_m &= \frac{\ln |z_m|}{T} \\ \omega_m &= \frac{\tan^{-1} \frac{Im(z_m)}{Re(z_m)}}{T}, \end{aligned} \tag{8.12}$$

where ln denotes the natural logarithmic function, and Re, Im denote the real and imaginary parts. The roots z_m are also used in Eqn. 8.7 to solve for h_m (using linear least square procedure), which in turn is used to calculate A_m and θ_m where

$$\begin{aligned} A_m &= |h_m| \\ \theta_m &= \tan^{-1} \frac{Im(h_m)}{Re(h_m)}. \end{aligned} \tag{8.13}$$

Thus all the parameters in the expansion of Eqn. 8.6 have been found.

8.4 BOXCAR ESTIMATION METHOD

The aim of most imaging problems is to find the image $f(x)$ from the collected data $G(k)$ according to the following integral equation

$$G(k) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi kx} dx. \tag{8.14}$$

If $G(k)$ is collected over an infinite frequency range, $f(x)$ can then be reconstructed accurately. However, in practice only a finite number of data $G(n\Delta k)$ where $n = 0, \ldots, N-1$, are measured which can be used to solve for an approximate value of $f(x)$ according to

$$f(x) = \Delta k \sum_{n=0}^{N-1} G(n\Delta k) e^{i2\pi n\Delta kx}. \tag{8.15}$$

Due to the finite sampling, $f(x)$ suffers from Gibb's oscillation and limited resolution. In 1989 Liang and Haacke [Liang *et al.* 1989] proposed a method using local polynomial approximation (LPA) of the image in order to overcome the drawbacks of finite sampling. Assuming $f(x)$ is spatially bounded, the LPA method divides $f(x)$ into M regions, with edges $[\epsilon_m, \epsilon_{m+1}]$, for $m = 1, \ldots, M$. In each region $f(x)$ is approximated by a polynomial with a certain order. The LPA of $f(x)$ can be written as [Liang *et al.* 1988, Boada *et al.* 1998]

$$f(x) = \sum_{m=1}^{M} (\sum_{r=0}^{R_m} C_{mr} (x - \beta_m)^r) \Pi_m(x), \tag{8.16}$$

where

$$\Pi_m(x) = \begin{cases} 1 & \text{for} \quad \epsilon_m < x < \epsilon_{m+1}, \\ 0 & \text{otherwise}, \end{cases} \tag{8.17}$$

$\beta_m = \frac{\epsilon_m + \epsilon_{m+1}}{2}$ and R_m, C_{mr} are the local polynomial order and coefficients respectively. Only the case of $R_m = 0$ will be discussed here, which reduces the LPA into a series of

boxcar functions with variable widths, amplitude and locations [Haacke *et al.* 1989a, Haacke *et al.* 1989b]. This model is useful in representing images that are piecewise uniform. By increasing the number of boxcar functions, smooth images can also be represented accurately. For the case $R_m = 0$, Eqn. 8.16 reduces to

$$f(x) = \sum_{m=1}^{M} C_m \Pi_m(x), \tag{8.18}$$

where M is the number of boxcar functions used to model $f(x)$. Using the fact that the Fourier transform of $\Pi_m(x) = d_m sinc(\pi d_m k) e^{-i2\pi\beta_m k}$, where $d_m = \epsilon_{m+1} - \epsilon_m$, Eqn. 8.14 can be written as

$$G(k) = \sum_{m=1}^{M} C_m d_m sinc(\pi d_m k) e^{-i2\pi\beta_m k}. \tag{8.19}$$

In order to reconstruct $f(x)$, the parameters d_m, β_m and C_m need to be found. Solving for these parameters from Eqn. 8.19 is a highly non-linear problem, which is time consuming to solve. If the method is changed to an all-pole model, then linear prediction theory can then be used to solve for the parameters. This is achieved by considering the first derivative of $f(x)$ where

$$\frac{df(x)}{dx} = \sum_{m=1}^{M+1} C'_m \delta(x - \epsilon_m), \tag{8.20}$$

where C'_m is the amplitude of $\delta(x - \epsilon_m)$. Since differentiating $f(x)$ in the space domain corresponds to multiplying the spectrum $G(k)$ by $i2\pi k$, we have

$$\hat{G}(k) = i2\pi k G(k) = \sum_{m=1}^{M'} C'_m e^{-i2\pi\epsilon_m k}, \tag{8.21}$$

where $M' = M + 1$. For simplicity, let $\hat{G}(k) = \hat{G}(n\Delta k) = \hat{G}(n)$, therefore Eqn. 8.21 can be written as

$$\begin{aligned} \hat{G}(n) &= i2\pi n\Delta k G(n) = \sum_{m=1}^{M'} C'_m e^{-i2\pi\epsilon_m n\Delta k} \\ &= \sum_{m=1}^{M'} C'_m z_m^n, \end{aligned} \tag{8.22}$$

where $z_m = e^{-i2\pi\epsilon_m \Delta k}$. Comparing Eqn. 8.22 to Eqn. 8.7, Prony's method can now be applied with $P = M'$ leading to the following linear prediction equations

$$\begin{Bmatrix} \hat{G}(2) & \hat{G}(3) & \dots & \hat{G}(M'+1) \\ \hat{G}(3) & \hat{G}(4) & \dots & \hat{G}(M'+2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{G}(N-M'+1) & \hat{G}(N-M'+2) & \dots & \hat{G}(N) \end{Bmatrix} \begin{Bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(M') \end{Bmatrix} = - \begin{Bmatrix} \hat{G}(1) \\ \hat{G}(2) \\ \vdots \\ \hat{G}(M') \end{Bmatrix} \quad (8.23)$$

Note that the minimum value of N is $2M'$, since the total number of unknowns is at least $2M+1$. As explained in section 8.3, once the ρ's are found by using an SVD based least squares procedure (for noise reduction and numerical stability), the roots z_m are the found by applying a polynomial factoring method which in turn are used to find the ϵ's by applying

$$\epsilon_m = \frac{-Im[\ln(z_m)] mod(2\pi)}{2\pi\Delta k}, \quad (8.24)$$

where $x \mod (y)$ is simply the remainder obtained from dividing x by y. The d's and β's are then calculated directly according to

$$\begin{aligned} d_m &= \epsilon_{m+1} - \epsilon_m \\ \beta_m &= \frac{\epsilon_{m+1} + \epsilon_m}{2}. \end{aligned} \quad (8.25)$$

The boxcar amplitudes, C_m, can be determined from Eqn. 8.19 such that

$$\vec{C} = (H^{T^*} H)^{-1} H^{T^*} \vec{G}, \quad (8.26)$$

where $H = d_m sinc(\pi d_m k) e^{-i2\pi\beta_m k}$, $\vec{G} = [G(0)G(1)\dots G(N-1)]^T$ and $\vec{C} = [C_1 C_2 \dots C_M]^T$. Once all the parameters are found, they can be substituted into Eqn. 8.18 to reconstruct $f(x)$.

The boxcar estimation method was studied by applying it on bandlimited version of a 1-D signal referred to as the high resolution signal. The bandlimited signal was obtained by the truncating the spectrum of the high resolution signal to the lowest 21 frequencies and then applying the inverse Fourier transform. Both the high resolution and bandlimited signals are shown in Figs. 8.1(a) and (b) respectively. The boxcar estimation method was applied on the spectrum of the bandlimited signal, and the reconstructed signals with $M = 1, 5, 12, 70$ are shown (in red) in Figs. 8.2(a),(b),(c) and (d) respectively. For $M = 1$ and 5, the number of boxcar functions are insufficient to model the bandlimited signal. For $M = 12$, the reconstructed signal has sharp edges and is a very good approximation of the main features of the bandlimited signal and free from Gibb's oscillations. However as M is increased to 70, the reconstructed signal approximates the bandlimited signal including Gibb's oscillation. Therefore by a

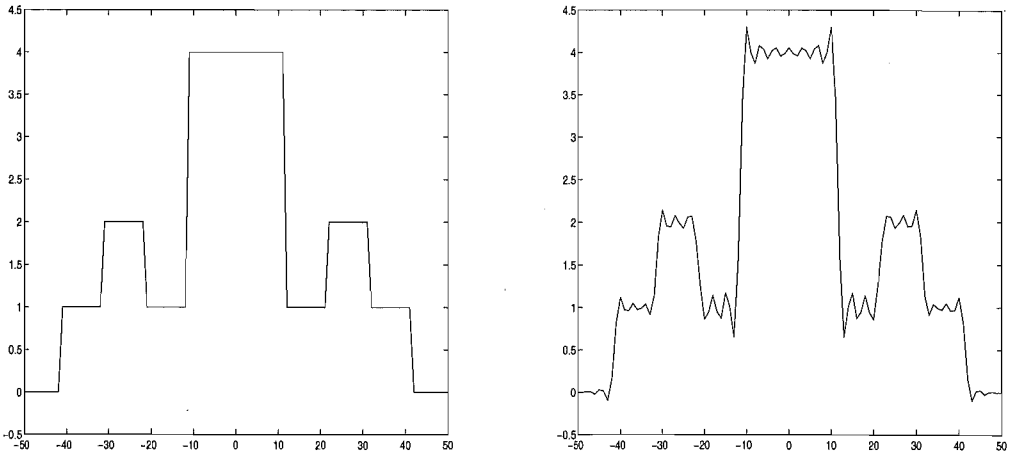


Figure 8.1 (a) High resolution signal; (b) bandlimited signal.

careful control of M, one is able to reconstruct a signal with sharp edges and accurately approximate the desired signal without the unwanted Gibb's oscillations.

Since the boxcar estimation method is inherently a 1-D technique, it could therefore not be directly applied to the bandlimited noisy head image, shown in Fig. 3.4(b). Following Haacke et al. [Haacke *et al.* 1989a] it was applied in one direction only, as follows. First, the inverse Fourier transform was applied along only the rows of the spectrum of the bandlimited head image (obtained by truncating the high resolution image spectrum to the lowest 101 spatial frequencies vertically and the lowest 201 frequencies horizontally). Then the estimation method was applied on each of the columns of the results independently. The reconstructed signals were then grouped back together to form the reconstructed image. Figs. 8.3(a) and (b) show the reconstructed images for $M = 15$ and $M = 40$ respectively. For $M = 15$, it is clear that the number of boxcar functions is insufficient to accurately approximate the original image, whereas for $M = 40$ seems that the reconstruction image sufficiently approximates the original image. Fig. 8.4 shows the ISNR curve obtained from applying the estimation method along the columns, which shows a maximum around $M = 40$.

Next, we applied the boxcar estimation technique (independently, in sequence) to both the rows and columns of the noisy bandlimited head image, with $M = 40$ in both cases. The resulting image is shown in Fig. 8.5. Since the estimation method is applied on each column and row independently, without taking into consideration any relationship that exists between neighbouring pixels, the reconstructed image edges are not smooth or continuous, however there is significant reduction of Gibb's oscillation in the reconstructed image. Figs. 8.6(a), (b), and (c) show profiles along a specific column of the high resolution head image, the noisy bandlimited head image and the reconstructed image obtained from applying the boxcar estimation method with $M = 40$ along both the rows and columns respectively. It is obvious the profile of the

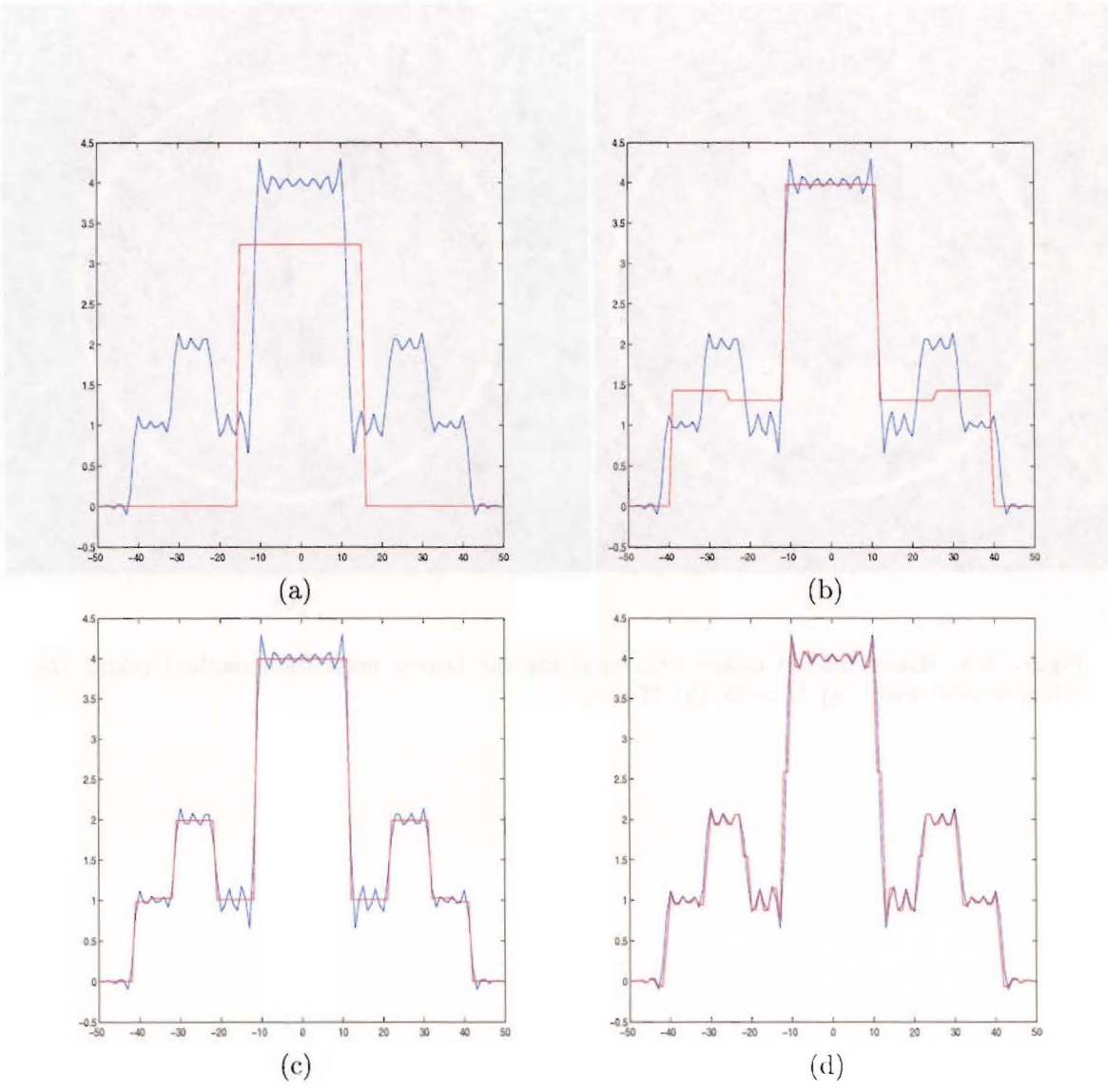
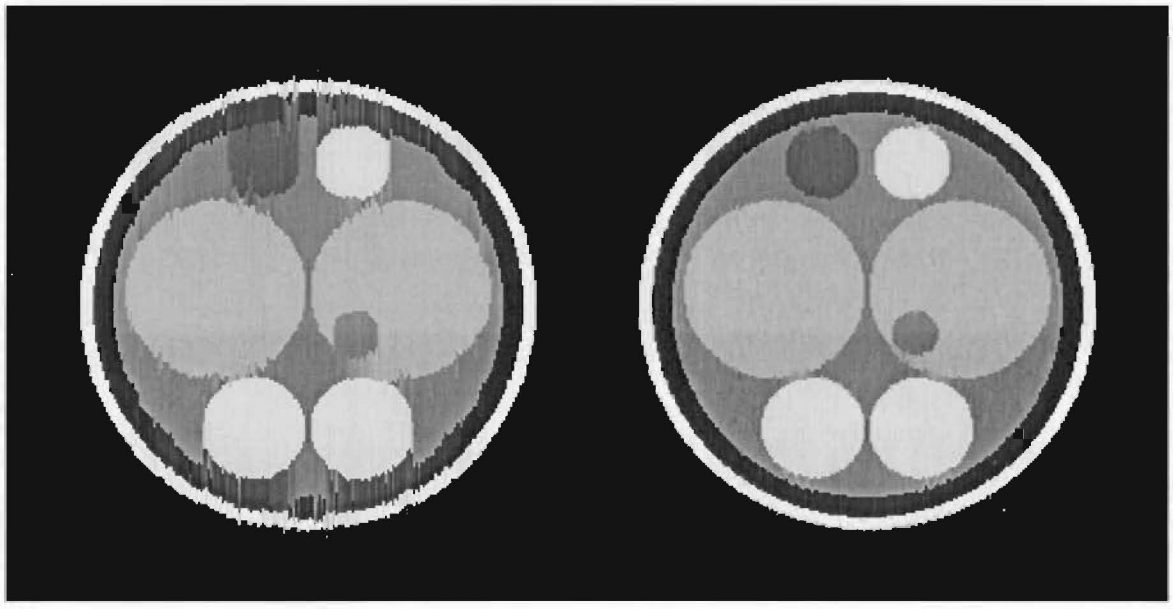


Figure 8.2 Boxcar estimation method where bandlimited signal is shown in blue and reconstructed signal shown in red for: (a)$M = 1$; (b) $M = 5$; (c)$M = 12$; (d)$M = 70$.



(a) (b)

Figure 8.3 Reconstructed image from applying the boxcar estimation method (along the columns only) with: (a) $M = 15$; (b) $M = 40$.

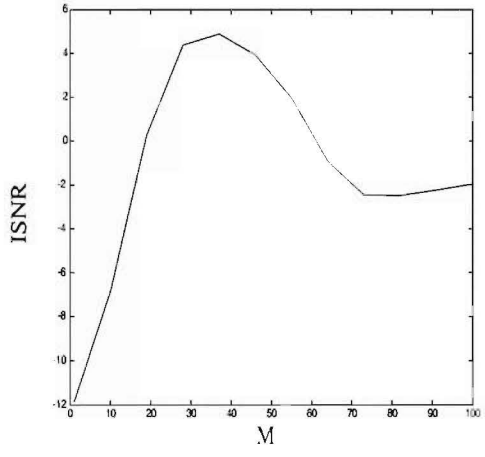


Figure 8.4 ISNR obtained from applying the boxcar estimation method (along columns only).

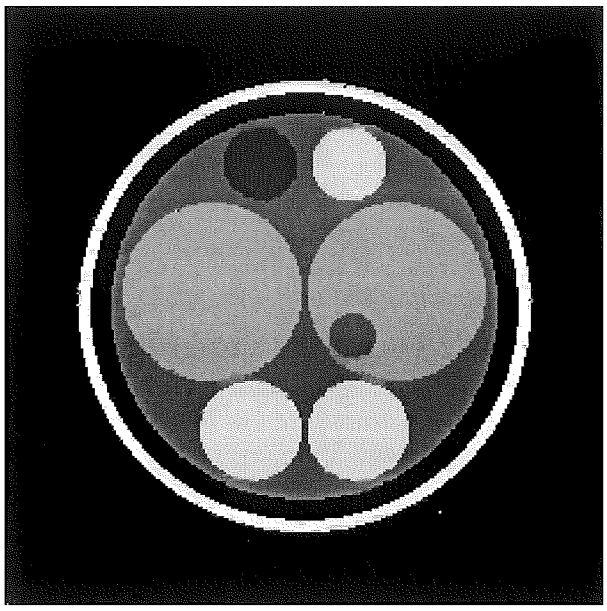


Figure 8.5 Reconstructed image from applying the boxcar estimation method (along both the columns and rows) with $M = 40$.

reconstructed image has sharper edges than those of the bandlimited image and less Gibb's oscillation.

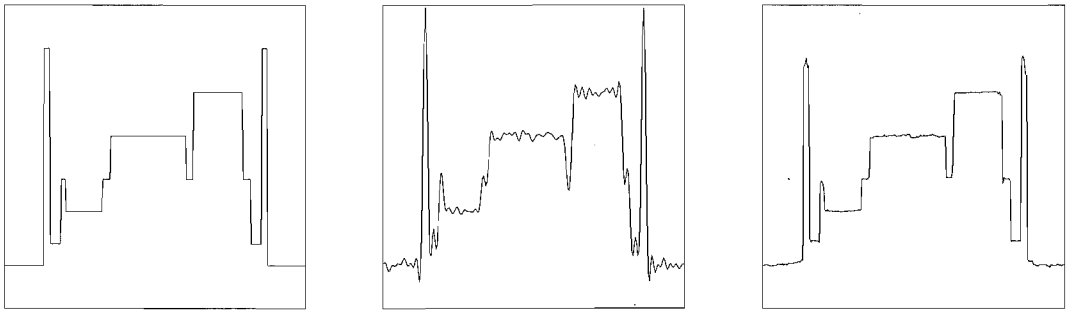
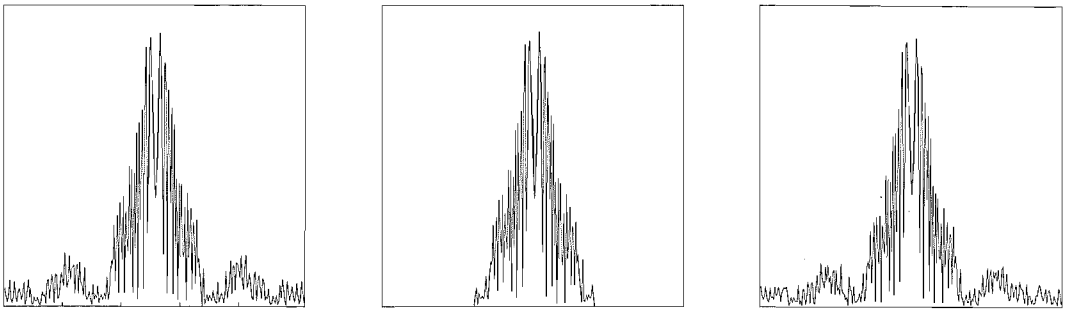


Figure 8.6 Profiles along: (a) high resolution head image; (b) bandlimited head image; (c) reconstructed image from applying estimation method with $M = 40$ along both the columns and rows.

Since the boxcar estimation method is a superresolving technique, the spectra of the reconstructed image beyond the imposed limited bandwidth should approximate that of the high resolution head image. Figs. 8.7(a), (b), and (c) show the magnitude profile along a particular column of the spectra of the high resolution image, the noisy bandlimited head image, and the reconstructed image obtained from applying the boxcar estimation method with $M = 40$ along both the rows and columns respectively. It can be clearly seen that the estimation method has recovered some of the spatial frequency information beyond the spatial frequency cutoff.



(a) (b) (c)

Figure 8.7 Profiles along a particular column of the spectra of: (a) high resolution head image; (b) bandlimited head image; (c) reconstructed image from applying estimation method with $M = 40$ along both the columns and rows.

8.5 GENERALIZED SERIES MODEL

A more general form of the parameter estimation method, known as the generalized series model, is presented here for completeness. Consider the image function $f(x)$ is represented as

$$f(x) = \sum_{n=1}^{p} c_n \Phi_n(x), \tag{8.27}$$

where c_n are series coefficients and $\Phi_n(x)$ are basis functions. Selection of the basis functions is the most important step in this technique since it determines how closely the model will approximate the image function. Liang and Lauterbur proposed weighted complex sinusoids [Liang and Lauterbur 1996] where

$$\begin{aligned} \Phi_n(x) &= C(x) e^{i2\pi n \Delta k x} \\ \text{and therefore} \quad f(x) &= C(x) \sum_{n=1}^{p} c_n e^{i2\pi n \Delta k x}. \end{aligned} \tag{8.28}$$

By choosing $C(x)$ to be a sum of boxcar functions, the generalized series models will comprise of two parts: the series of boxcar functions (useful for modelling sharp edges) and the Fourier series model (useful for modelling smooth functions), therefore enabling it to represent different features in the image function. It is also been proposed to use the generalized series model for dynamic imaging to reconstruct high resolution dynamic images from sparsely sampled data where a full acquired data set of a reference image is used in determining the basis functions [Liang and Lauterbur 1994].

There are many other proposed parameter estimation methods, such as the autoregressive moving average model (ARMA) where a image function $f(x)$ is modelled by a

rational function such that

$$\begin{aligned} f(x) &= \frac{\sum_{n=0}^{q} b_n e^{i2\pi n\Delta kx}}{1+\sum_{n=1}^{p} a_n e^{i2\pi n\Delta kx}} \\ &= \frac{b_0 + b_1 z + \ldots + b_q z^q}{1 + a_1 z + \ldots + a_p z^p} \\ \text{where } z &= e^{i2\pi\Delta kx}, \end{aligned} \tag{8.29}$$

and the parameters p, q, a_n and b_n need to be solved. Note that when $p = 0$, the ARMA model is simply the Fourier series model, which is useful for representing smooth functions but is not as good when approximating spiky functions, whereas the ARMA model is useful for spiky functions. There are different techniques to solving the parameters of the ARMA model, such as the transient error method discussed by Smith et al. [Henkelman and J.Bronskill 1987].

8.6 NONCONVEX LEVEL PENALTY FUNCTION

A new constraint, known as the level penalty function and based on the prior knowledge of the possible M likely levels that can be attained by the pixels, is introduced in this section. The constraint aims to find an image (a solution) with a histogram that approximates the histogram of the high resolution image. This is achieved by designing the constraint to penalize pixels at unlikely levels [Connolly *et al.* 2000]. The penalty function chosen is simply the natural logarithm of a sum of Gaussian functions approximating the expected image's histogram and can be written as [Alwesh *et al.* 1998]

$$\begin{aligned} \mathcal{S}(f) &= -\sum \ln\left(p(f)\right) \\ &= -\sum \ln\left(\chi_1 e^{-\beta_1(f-\gamma_1)^2} + \chi_2 e^{-\beta_2(f-\gamma_2)^2} + \ldots + \chi_M e^{-\beta_M(f-\gamma_M)^2}\right) \end{aligned} \tag{8.30}$$

where $p(f)$ is the probability density function and the summation is over all pixels. The roles of the different variables χ, β, γ can be understood from Fig. 8.8 which shows a plot of $p(f)$ ($\chi_1 > \chi_M$, $\beta_1 > \beta_2$, $\gamma_1 < \gamma_2$), where it is obvious that χ controls the height of the Gaussian function, β controls the width and γ controls the location (mean) of the Gaussian function. By changing these variables, different approximations to the desired histogram can be obtained based on some prior knowledge of the levels present in the original image.

Differentiating $\mathcal{S}(f)$ with respect to f then substituting the derivative into Eqn.

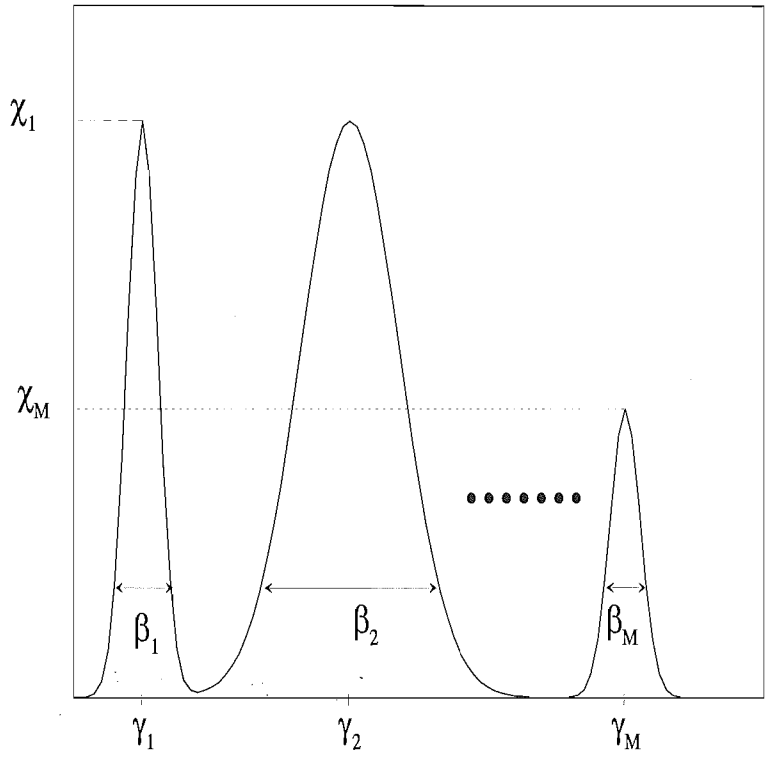


Figure 8.8 Illustration of the effect of the χ, β and γ variables on the shape of $p(f)$.

5.23 we have

$$\mathcal{S}'(f) = \frac{2\chi_1\beta_1(f-\gamma_1)e^{-\beta_1(f-\gamma_1)^2} + \ldots + 2\chi_M\beta_M(f-\gamma_M)e^{-\beta_M(f-\gamma_M)^2}}{\chi_1 e^{-\beta_1(f-\gamma_1)^2} + \ldots + \chi_M e^{-\beta_M(f-\gamma_M)^2}}$$

$$f^{k+1} = f^k + 2\alpha^k H^* G - 2\alpha^k H^* H f^k - \alpha^k \lambda \mathcal{S}'(f). \tag{8.31}$$

It should be noted that with a single level (i.e. $p(f)$ in Eqn. 8.30 is then a single Gaussian function), Eqn. 8.31 reduces to a Tikhonov regularization with a single level as a model function (discussed in section 5.4). Two different sets of sums of Gaussian functions $(p_1(f), p_2(f))$ were studied for the head data. Figs. 8.9 (a) and (b) show the histogram of the high resolution and the bandlimited head image within the support region respectively. Note that there are seven distinct levels ($[0, 0.1, 0.25, 0.4, 0.6, 0.8, 1]$) in the high resolution head image.

Fig. 8.10 shows the function $p_1(f)$ for γ_i values $[0.1, 0.25, 0.4, 0.6, 0.8, 1]$ (upper figure) and the corresponding gradient function (lower figure). Note that $\mathcal{S}'$ at pixels with the chosen γ_i values is zero, indicating that no penalty is applied to pixels with these values and the farther the pixels' values deviate from the chosen γ_i values, the stronger is the applied penalty. The nonconvex level penalty function was applied on the bandlimited head image with $p_1(f)$ as the approximation to the expected image's histogram and $\lambda = 0.0001$.

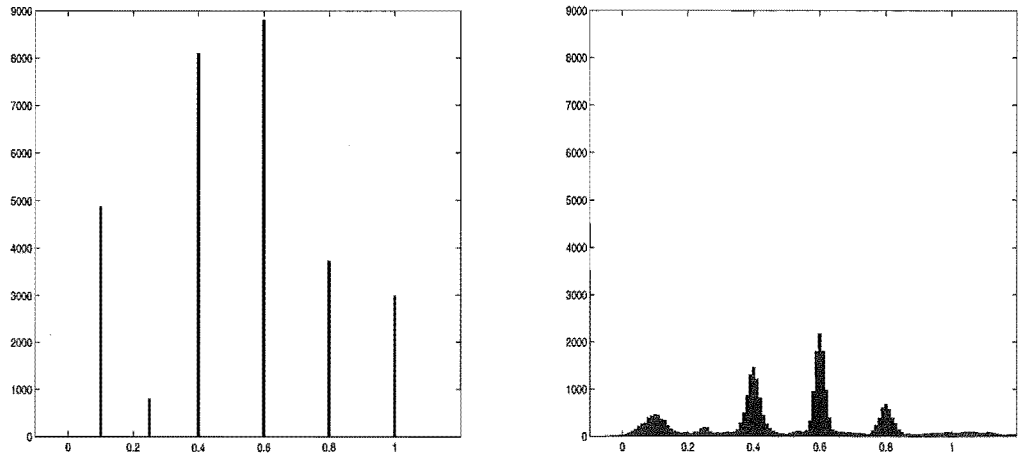


Figure 8.9 Histograms of: (a) high resolution head image; (b) bandlimited noisy head image.

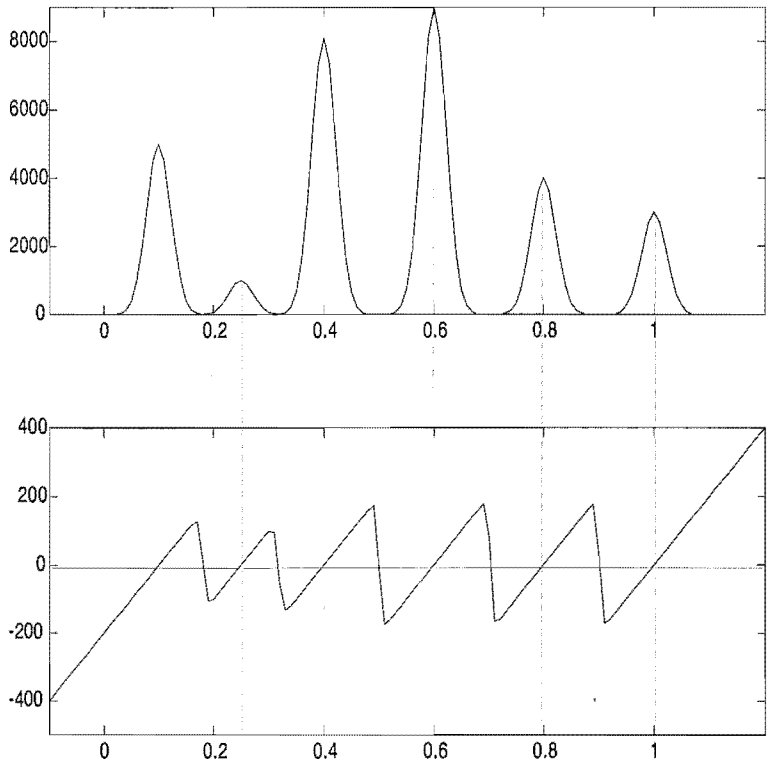
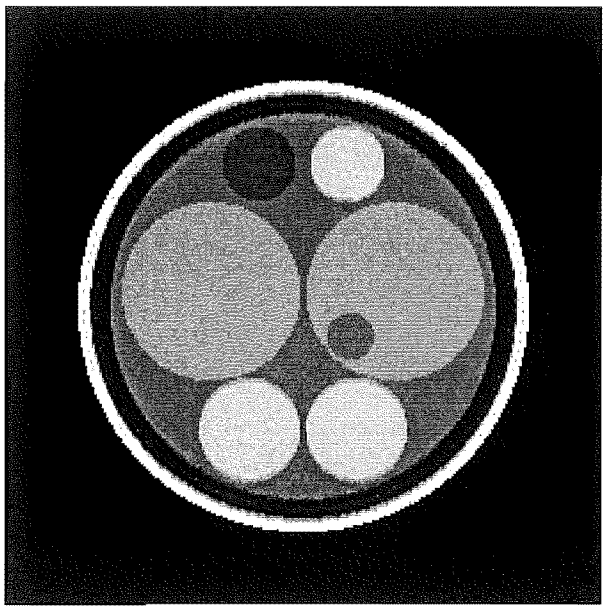
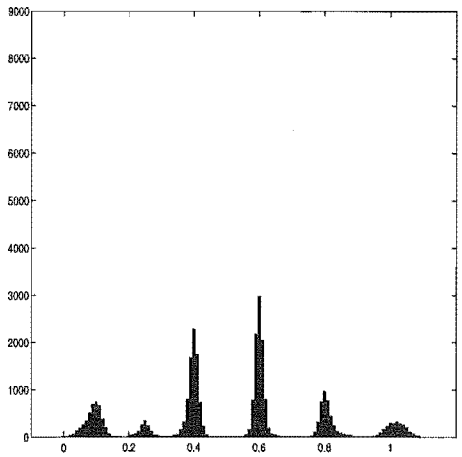


Figure 8.10 Illustration of the function $p_1(f)$ and the corresponding $\mathcal{S}'(f)$.

Figs. 8.11(a) and (b) show the recovered image (at iteration 1000) and the corresponding histogram. The recovered image has some sharper edges, less Gibb's oscillation than that of the bandlimited head image, and the corresponding histogram is closer to the histogram of the high resolution image than the histogram of the bandlimited image. The corresponding ISNR curve, shown in Fig. 8.14, has a higher peak than that of the G-P algorithm and is regularized. The exact same results were obtained when $\chi_1 = \chi_2 = \ldots = \chi_6$, thereby reducing the amount of prior information needed i.e. it is not necessary to know the exact histogram of the solution to implement this method. Further improvements to the solution might be expected to be obtained with an additional assumption of local smoothness. This could be implemented by combining the penalty with an additional penalty function such as the truncated quadratic penalty, discussed in section 5.6.3. Note such an extension is similar to improving the multiple level quantization with the neighbourhood based quantization, discussed in section 4.9.2.



(a)



(b)

Figure 8.11 Nonconvex level penalty function using $p_1(f)$ and $\lambda = 0.0001$: (a) recovered head image; (b) histogram.

The second penalty function $p_2(f)$, shown in Fig. 8.12, was deliberately designed to strongly penalize pixels at level 0.6. This was done to study the case when a correct level (that exists in the high resolution image) is wrongly assumed not to exist. Fig. 8.13(a) shows the recovered image (at iteration 1000) when the penalty function $p_2(f)$ was used. In this case, pixels with values around 0.6 (located in the two big circles), were set to values close to either 0.4 or 0.8. This is confirmed in the corresponding histogram shown in Fig. 8.13(b) which indicates the absence of pixels at level 0.6 and the deteriorated ISNR curve shown in Fig. 8.14. Therefore care must be taken when

designing the approximation to the desired histogram, since the absence of correct levels will adversely affect the recovered image.

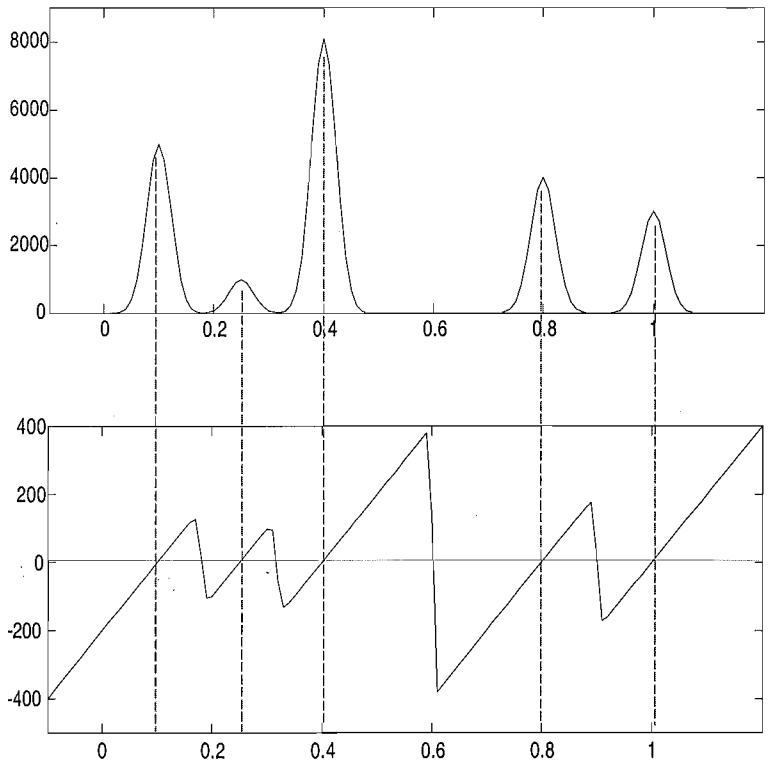


Figure 8.12 Illustration of the function $p_2(f)$ and the corresponding $\mathcal{S}'(f)$.

The nonconvex level penalty function was also applied on the bandlimited noisy brain image. Fig. 8.15(a) shows the histogram of the high resolution brain image (within the support region) and the $p_3(f)$ function (the estimate of the desired histogram in red). Fig. 8.15(b) shows the recovered image with $\lambda = 5$. Note the undesirable blocky appearance and lack of recovery of additional detail, comparing Fig. 8.15(b) with Fig. 3.5(a), however it is visually preferable compared to the bandlimited brain image, shown in Fig. 3.5(b). The disappointing results probably reflect the relatively poor match of the piecewise uniform model to the original.

It is important to note that applying the nonconvex level penalty function might result in a local minima, but this problem was not faced when applied to the head and brain data since the algorithm was started with a good approximation to the true solution.

8.7 CONCLUSION

An alternative to the conventional method of reconstructing MR images is the modelling of data based on some prior knowledge of the nature of the imaged object (i.e. spiky, or sharp edges or smooth). One of the problems of data modelling methods is the need

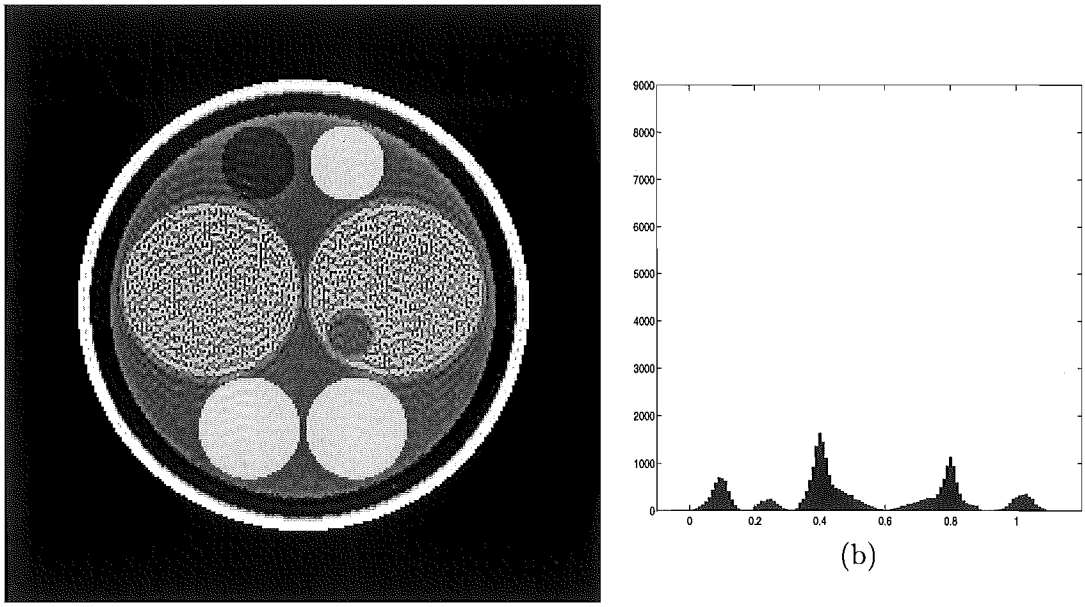


Figure 8.13 Nonconvex level penalty function using $p_2(f)$ and $\lambda = 0.0001$: (a) recovered head image; (b) histogram.

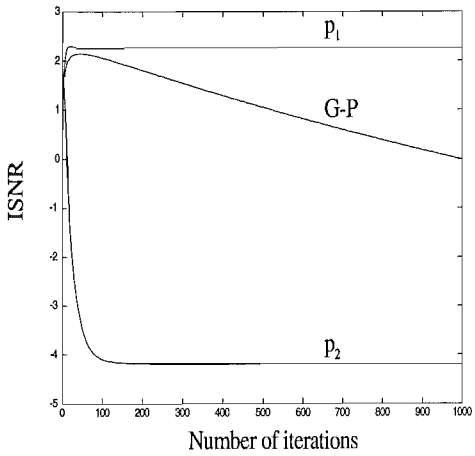
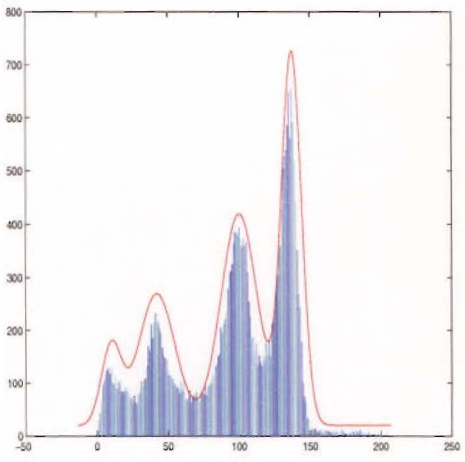
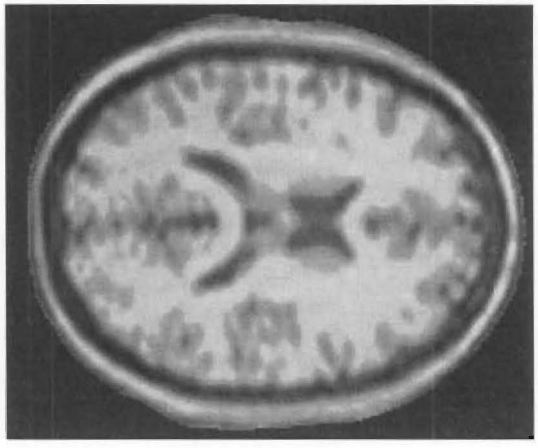


Figure 8.14 ISNR curves obtained from applying the nonconvex level penalty functions $p_1(f)$ and $p_2(f)$.



(a)



(b)

Figure 8.15 (a) Histogram of high resolution brain image in blue, $p_3(f)$ in red; (b) recovered image using nonconvex level penalty function.

to determine different parameters such as the polynomial order, the number of boxcar functions and basis functions which affect how closely the model approximates the function. The boxcar estimation method is adequate for representing image functions which are piecewise uniform. If the number of boxcar functions M is carefully chosen, the image can be almost exactly approximated without Gibb's oscillations. If M is chosen too large, the reconstructed image tends to suffer from Gibb's oscillation or else the system of linear prediction equations becomes unstable. In the latter case, the SVD method with small valued eigenvalues set to zero needs to be applied. Another major drawback is that the boxcar estimation method is inherently a 1-D technique, therefore when applied on a 2-D image (by applying it independently on each row and column), the edges of the image are not continuous or smooth.

Another method that utilizes prior information in the reconstruction process, is the nonconvex level penalty function. This method is based on some knowledge of the shape of the desired histogram. It penalizes pixels with specific level values to obtain an image with a histogram that approximates that of the high resolution image. Application of this method on piecewise uniform images results in images with sharp edges and less Gibb's oscillations. However, results on a more realistic image are disappointing.

Chapter 9

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This thesis addressed the problem of superresolution specifically in magnetic resonance (MR) imaging. Different existing and new methods were studied by applying them both on a phantom and realistic MR data. The well-known Gerchberg-Papoulis (G-P) algorithm, when applied on both types of available data, resulted in minimal visual recovery. Therefore, additional constraints were incorporated into the G-P algorithm such as positivity and total energy, as proposed previously, and the new single level and neighbourhood based constraints. Both visual study of the recovered image and the improvement in the signal-to-noise ratio (ISNR) were used as criteria to measure the success of the applied algorithm. In most cases the ISNR curve was found to diverge due to the ill-posed nature of the problem indicating the need for regularization. In particular, the total variation method provided edges preserving regularization with sharp images obtained when it was applied to piecewise constant images. The extension of the superresolution problem to undersampled images provided important results where significant and sometimes even complete recovery was achieved using some undersampling patterns. This thesis also addressed the problem of superresolving a region of interest (ROI), which is important in dynamic imaging applications and follow up studies. Existing methods, such as the reduced field of view method, were compared to the new iterative region of interest (iROI) method which provided significant recovery of the ROI and which required less prior information regarding the object when using another version of the algorithm. Another method that utilizes prior information, the nonconvex level penalty function, was presented and which resulted in sharper images when applied to piecewise homogeneous images. The following paragraphs provide more detailed comments related to specific chapters of the thesis.

The G-P algorithm has been extended in the work presented here to incorporate more prior information in superresolving MR images. As originally proposed, the G-P algorithm incorporated only a support constraint in the spatial domain. The basic G-P algorithm was applied on both the phantom and MR brain data, and the results, presented in section 3.7, show minimal visual improvement for both types of data,

indicating the need for additional constraints to be applied. In this work, when the G-P algorithm or modifications of the G-P algorithm were applied, the basic steepest descent form was utilized, since the major focus of this study was on the success of recovery rather than efficiency. In future work, the conjugate gradient method should be utilized which reduces computation time considerably. Care should be taken however when applying the conjugate gradient method to problems that are ill-posed since though the method approaches the solution (i.e. the iteration with maximum ISNR) faster than the steepest descent algorithm, it also diverges faster away from this solution.

In addition to the prior information used in the basic G-P algorithm, other known properties of the object were incorporated into the problem by utilizing the projection onto convex sets (POCS) method, discussed in chapter 4. In this method, each known property is associated to a convex set and is used as a constraint to find a solution that satisfies all constraints (i.e. lies in the intersection of all the convex sets). The different constraints studied were the positivity, energy, and upper and lower bound constraints, all of which resulted in some improvement in the ISNR curves compared to that of the G-P algorithm. However, none provided more than limited visual improvement.

A more general form of the POCS method was then studied, known as the method of generalized projections (MGP), where nonconvex constraints may be incorporated into the problem, such as the single level quantization $\mathcal{P}_6$. This projection-based operator was shown to result in a sharper image, in regions containing the chosen level, however care was found to be necessary when choosing $\delta\mathcal{P}_6$, since large values of $\delta\mathcal{P}_6$ will result in pixels being set to the wrong level. Application of the single level quantization requires prior approximate knowledge of the levels present in the object. As mentioned in section 4.9.1, the single level quantization is a harsh constraint that does not take into account any relationship that might exist between neighbouring pixels. This drawback was then remedied by using the neighbourhood based quantization ($\mathcal{P}_7$) which sets a pixel value to a prescribed level if the majority of pixels surrounding the pixel under study have values within a specific deviation from the prescribed level. This operator is based on the assumption that neighbouring pixels belong to the same homogeneous tissue and when applied to the piecewise continuous head phantom provided significant visual improvement. Combinations of the above constraints were also studied with the aim of obtaining a levelled off ISNR curve with a higher maximum than those obtained when applying the constraints independently. In this thesis only three combination sets were investigated. A more thorough study of the effect of changing the orders of applying several constraints and the effect of relaxing the constraints should be part of future research.

One of the problems of the G-P algorithm is that it is sensitive to noise due to the ill-posed nature of the problem (i.e. the solution does not continuously depend on the data). In this case regularization methods (discussed in chapter 5) are required

to stabilize the problem. There are several different regularization methods such as the Tikhonov-Miller regularization method (when no prior information regarding the smoothness of the object is available) which stabilizes a problem by incorporating a penalty function which represents a bound on the energy of the image. The Tikhonov-Miller regularization was applied on both the phantom and the brain data, and in both cases the recovered images obtained at later iterations were very similar to the corresponding recovered images obtained from applying the G-P algorithm (at the iteration with the maximum ISNR value), indicating a stable system with a levelled off ISNR.

Another form of the Tikhonov-Miller regularization is the model-based regularization, studied in section 5.4, where a model of the solution is incorporated into the algorithm. This method is quite useful when applied to cases where high and low resolution data are alternatively acquired of adjacent slices. In order to superresolve the low resolution slice, the model used is simply the average of the high resolution slices sandwiching the low resolution slice. The closer the slices are to each other, the closer the model used is to the true solution and the further the slices are from each other the more incorrect information is introduced into the solution. Therefore care must be taken in selecting the regularization parameter, since a small regularization parameter value will hardly result in any regularization and a large regularization parameter value will force the solution to be the same as the incorporated model. This regularization method was applied to the brain data, where the model used contained both some similarities and some differences to the true solution, and it provided a considerable recovery of the bandlimited brain slice.

Up to this point, information regarding the smoothness of the image was assumed to be unavailable. In cases where the object is known to be smooth, constraints can be used which force neighbouring pixels to have similar intensity values and provide a regularized solution. The effect of imposing a smoothness constraint was quite obvious from the recovered image (presented in section 5.5) where the edges were smoothed out and Gibb's oscillations reduced.

The total variation method is an edge preserving regularization method that aims to find a solution with the least total variation and is thus suitable for blocky images which have a low total variation. It provided significant edge enhancement when applied to bandlimited head data and cleaned up most of the Gibb's oscillation (which has a relatively high total variation). However, when applied on the bandlimited brain data, it resulted in an undesirable blocky image. Future research should involve testing of the total variation penalty on real images that are actually blocky (i.e. reasonably close to piecewise constant) in nature.

A compromise between smoothness and edge preservation can be achieved by using the Huber function which applies the smoothness constraint for gradient values less than

a chosen gradient value (M) and total variation for gradient values greater than M. By varying M and the regularization parameter, different amounts of smoothness and edge preservation can be achieved. For most of the regularization methods studied, the steplength had to be reduced to avoid the production of artifacts in the image due to the nonconvergence of the algorithm.

In this thesis, superresolution was not limited to the bandlimited case but was extended to the undersampling of the spatial frequencies. Different undersampling patterns were studied in chapter 6 by applying them on both the head and brain data. The G-P algorithm with the positivity constraint was then applied to the undersampled spatial frequency data. In some cases there was significant (and even complete) recovery of aliased regions even when the aliasing seemed severe. The amount of recovery seemed to depend on two factors 1) determinacy and 2) the shape of the undersampling pattern. When the undersampling pattern was such that a missed out spectral component was surrounded by retained components and the problem was overdetermined, significant and sometimes even complete recovery was achieved. For cases where the problem was underdetermined and the undersampling pattern used had missing components surrounded by other missing components, minimal (or no) recovery was achieved. However we note that for other undersampling patterns with missing components surrounded by retained components considerable recovery was achieved even if it was an underdetermined system.

Since the above method is not an ad hoc method and the formulation of the aliased image in terms of the original image was possible, future work should be directed towards formulating the recovery image in terms of the original image. Therefore if only a region of the image is of interest, a specific undersampling pattern could be chosen based on the prior knowledge of which regions of the image it will clean from aliasing. In addition, by formulating the aliased image in terms of the original image (as was done in section 6.6), knowledge of the regions that will be aliased is obtained. In this case, the unaliased regions can be directly used (without applying any superresolution algorithms) and prior knowledge on the solution (such as the constraints considered in this thesis) can be used to improve recovery within the aliased regions only.

For some applications such as dynamic imaging or follow up of a tumour, changes are limited to a region of interest (ROI) and acquiring high resolution data in every session will contain substantial redundancy. There are several methods available which can be used for such applications to reduce the imaging time. The keyhole method, discussed in section 7.3, is a very simple method which sets the unmeasured high spatial frequency data of an acquired low resolution dynamic image (e.g. big tumour image) to the corresponding high spatial frequency data of a high resolution reference image (e.g. small tumour image). Results obtained from applying this method to the MR brain data show a sharper image but with very little extra information compared to the conventional bandlimited image.

For cases where the changes in a dynamic session are limited to a small area of the entire field of view, the reduced field of view method can be used in such applications since it reduces imaging time and provides high resolution dynamic images. It requires some approximate prior knowledge of the extent and position of the ROI to determine the reduction factor and help in repositioning the ROI. The major disadvantage of this method is that any change occurring outside the reduced field of view will be aliased into the reduced field of view, therefore it cannot be used for some applications such as cardiac and abdominal imaging.

To overcome the above drawback a new method with two versions, known as the iterative recovery of ROI (iROI) was presented in section 7.5. For the first version of this method, a high resolution reference image is incorporated into the G-P algorithm to superresolve a ROI of a low resolution dynamic image. Similar to the rFOV method, this method requires some approximate knowledge of the size and location of the ROI. Impressive recovery of the ROI was achieved when applying the iROI method with the exact mask of the ROI on the bandlimited dynamic brain data and significant recovery was obtained when an approximate mask and the total variation method were used. When using approximate masks and the total variation method, even though the recovered tumour edges were sharp, the recovered image within the mask suffered from a ringing artifact which did not seem to compromise information regarding the tumour. Future work should involve reducing or even eliminating this artifact. An easy to implement and fast way of reducing the ringing is by using a weighting function for the amount of regularization λ such that λ is high at the centre of the mask and slowly decreases as it approaches the edges of the mask. This requires designing a weighting function that might have a flat part in the center (where edge-preservation is very important) and that slowly rolls off to zero as it approaches the edges. Another technique to reduce the ringing is by designing a penalty function which acts as the total variation regularization method in the center and as the smoothness constraint at the edges. The decision of applying either the total variation or smoothness constraint would then depend on the spatial location of the pixel under consideration. This method is more complicated since it would involve taking into consideration several parameters.

The iROI method was also applied on undersampled dynamic brain data using a high resolution reference data which resulted in an image with sharp edges and free from any aliasing. The second version of the iROI method with the total variation method did not provide results quite as good as the first version but has the major advantage that the reference image used here is a low resolution image thus requiring less prior knowledge.

In addition to the iterative methods, parametric estimation methods were also used to superresolve images by utilizing a model based on prior knowledge of the image. The boxcar estimation method, discussed in chapter 8, models a piecewise uniform

image by a series of boxcar functions, whose widths, amplitudes and locations need to be computed. The boxcar estimation method was applied independently on the rows and columns of the bandlimited head image resulting in a recovered image with sharp edges and reduced Gibb's oscillation. Care needs to be taken when choosing the number of boxcar functions, since a large number of boxcar functions will approximate the image including the Gibb's oscillations. It is important to note that the edges of the recovered image were not smooth or continuous, since no consideration was taken regarding any relationship that might exist between neighbouring pixels. To overcome this problem, future research should be directed towards designing an inherently 2-D boxcar estimation method which is applied on the whole of the 2-D image.

Another method presented that utilizes prior information in the superresolution process is the nonconvex level penalty function. It aims to find a solution with a histogram approximating the histogram of the true solution based on the prior knowledge of the levels present in the true solution. The nonconvex level penalty function resulted in a sharper image with less Gibb's oscillation and a levelled off ISNR curve (i.e. the problem is stable) when the correct levels were assumed. Again the brain image is too variable to be a good test here, a simple realistic image (e.g. leg) would be more suitable.

In conclusion, the G-P algorithm and its extensions (i.e. including the additional constraints) seemed to provide better results when applied on the bandlimited head image in comparison to that of the more realistic MR data, indicating the need for more prior information for the recovery of the brain image. The additional constraints may be either convex or nonconvex, such as the neighbourhood based method which utilizes prior knowledge regarding the levels present in the true solution in the recovery process. Another method which requires prior knowledge of the levels present is the nonconvex penalty function. Both the neighbourhood based and nonconvex level penalty function provide best results when applied to a piecewise homogeneous image.

It was also found that the superresolution problem is easier to solve when applied to undersampled images where limited prior information is required in the recovery process. In some cases complete removal of the aliasing was achieved by selecting a suitable undersampling pattern. Superresolution of a region of interest also seems easier than that of a whole image, which is useful in applications such as dynamic imaging and follow-up studies. In most cases, regularization methods are also important to overcome the ill-posedness nature of the problem. The total variation method, when applied to piecewise constant images, provided edge-preservation as well as regularization. The boxcar estimation in 1-D also provided sharp edges, however when it was applied on 2-D images the edges were not continuous since the estimation method was applied independently on each row and column of the image.

REFERENCES

AHN, C.B., KIM, J.H. AND CHO, Z.H. (1986), 'High-speed spiral-scan echo planar NMR imaging-I', *IEEE Transactions on Medical Imaging*, Vol. 5, pp. 2–7.

ALWESH, N., BONES, P.J. AND CONNOLLY, T.J. (1998), 'Superresolution of MR images', *Proc. Image and Vision Computing New Zealand, IVCNZ98*, November, pp. 246–251.

ALWESH, N., CONNOLLY, T.J. AND BONES, P.J. (1999), 'Recovery of undersampled magnetic resonance images', *Proc. Image and Vision Computing New Zealand, IVCNZ99*, August, pp. 265–270.

ANDREWS, H.C. AND HUNT, B.R. (1977), *Digital image restoration*, Prentice-Hall Signal Processing Series.

APOSTOL, T. (1958), *Mathematical analysis*, Addison-Wesley.

ATLAS, S.W. (1996), *Magnetic resonance imaging of the brain and spine*, Lippincott-Raven.

BAKER, C.T.H. AND MILLER, G.F. (1982), *Treatment of integral equations by numerical methods*, Academic Press.

BALLINGER, J.R. (1996), 'What is MRI', http://128.227.164.224/mritutor/WhatisMR.html.

BALTES, H.P. (Ed.) (1980), *Inverse scattering problems in optics*, Springer-Verlag.

BANHAM, M.R. AND KATSAGGELOS, A.K. (1997), 'Digital image restoration', Vol. 81, No. 2, March, pp. 24–41.

BARAKAT, V., GUILPART, B., GOUTTE, R. AND PROST, R. (1997), 'Model-based Tikhonov-Miller image restoration', *International Conference on Image Processing*, Vol. 1, pp. 310–311.

BLAKE, A. AND ZISSERMAN, A. (1987), 'Visual reconstruction', *MIT Press.*

BLOCH, F. (1946), 'Nuclear induction', *Physical Review*, Vol. 70, pp. 460–473.

BOADA, F., LIANG, Z.P. AND HAACKE, E.M. (1998), 'Improved parametric reconstruction using variable projection optimization', *Inverse Problems*, Vol. 14, pp. 19–27.

BONES, P.J., ALWESH, N. AND CONNOLLY, T.J. (1998), 'Superresolution of MR images', Proc. Summer Topical Meeting, Signal Recovery and Synthesis, Optical Society of America, Hawaii, pp. 120–122.

BORNERT, P., SCHOMBERG, H., ALDEFELD, B. AND GROEN, J. (1999), 'Improvements in spiral MR imaging', *Magnetic Resonance Materials in Physics, Biology and Medicine*, Vol. 9, pp. 29–41.

BOUMAN, C. AND SAUER, K. (1993), 'A generalized Gaussian image model for edge-preserving MAP estimation', *IEEE Transactions on Image Processing*, Vol. 2, No. 3, pp. 296–310.

BRACEWELL, R.N. (1995), *Two-dimensional imaging*, Prentice Hall.

CHAN, T. AND MULET, P. (1996), 'Iterative methods for total variation image restoration', *UCLA Computational and Applied Mathematics Reports*, pp. 1–24. (See also http://www.math.ucla.edu/applied/cam/index.html).

CHO, Z.H., JONES, J.P. AND SINGH, M. (1993), *Foundations of medical imaging*, John Wiley and Sons, Inc.

COLLINS, D.L., ZIJDENBOS, A.P., KOLLOKIAN, V., SLED, J.G., KABANI, N., HOLMES, C.J. AND EVANS, A.C. (1998), 'Design and construction of a realistic digital brain phantom', *IEEE Transactions on Medical Imaging*, Vol. 17, No. 3, pp. 463–468. (See also http://www.bic.mni.mcgill.ca/brainweb/).

COMBETTES, P.L. (1993), 'The foundation of set theoretic estimation', *Proceedings of the IEEE*, February, pp. 182–208.

COMBETTES, P.L. (1994), 'Set theoretic signal processing', *Proceedings of the Seventh IEEE Workshop on Statistical Signal and Array Processing*, pp. 1–6.

COMBETTES, P.L. AND CIVANLAR, M.R. (1991), 'The foundations of set theoretic estimation', *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 2921–2924.

COMBETTES, P.L. AND PUH, H. (1993), 'Parallel projection methods for set theoretic signal reonstruction and restoration', *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 15, pp. 297–300.

CONNOLLY, T.J. AND LANE, R.G. (1997), 'Gradient methods for superresolution', *Proceedings of IEEE International Conference on Image Processing XI*, Vol. 1, pp. 917–920.

CONNOLLY, T.J., LANDMAN, K.A. AND WHITE, L.R. (1995), 'On Gerchberg's method for the Fourier inverse problem', *Journal of Australian Mathematical Society, Series B*, Vol. 37, pp. 26–44.

CONNOLLY, T.J., ALWESH, N. AND BONES, P.J. (1999), 'Total variation based super-resolution', *Proc. Image and Vision Computing New Zealand, IVCNZ99*, August, pp. 283–288.

CONNOLLY, T.J., ALWESH, N. AND BONES, P.J. (2000), 'Utilizing prior knowledge for superresolution', Australia and N.Z. Industrial Applied Mathematics Conference, ANZIAM 2000, Copthorne Resort, Waitangi, Bay of Islands, NZ.

COSOSCO, C.A., KOLLOKIAN, V., KWAN, R.K.S. AND EVANS, A.C. (1997), 'Brain-web:online interface to a 3D MRI simulated brain database', *NeuroImage*, Vol. 5, No. 4. (See http://www.bic.mni.mcgill.ca/brainweb/.

DELANEY, A.H. AND BRESLER, Y. (1998), 'Globally convergent edge-preserving regularized construction: An application to limited-angle tomography', *IEEE Transactions on Image Processing*, Vol. 7, No. 2, February, pp. 204–221.

DEN DEKKER, A.J. AND VAN DEN BOS, A. (1997), 'Resolution: a survey', *Journal of Optical Society of America*, Vol. 14, No. 3, March, pp. 547–557.

DOBSON, D.C. AND SANTOSA, F. (1996), 'Recovery of blocky images from noisy and blurred data', *SIAM J. Sci. Comput.*, Vol. 56, No. 4, pp. 1181–1198.

DUYN, J.H. AND YANG, Y. (1997), 'Fast spiral magnetic resonance imaging with trapezoidal gradients', *Journal of Magnetic Resonance*, Vol. 128, pp. 130–134.

FEINBERG, D., HALE, J., WATTS, J., KAUFMAN, L. AND MARK, A. (1986), 'Halving MR imaging time by conjugation: demonstration at 3.5kG', *Radiology*, Vol. 161, pp. 527–531.

FELIX, W., WEHRLI, D.S. AND KNEELAND, J.B. (1988), *Biomedical magnetic resonance imaging, principles, methodology, and applications*, VCH Publishers.

GALATSANOS, N.P. AND KATSAGGELOS, A.K. (1992), 'Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation', *IEEE Transactions on Medical Imaging*, Vol. 1, No. 3, July, pp. 322–336.

GEMAN, D. AND REYNOLDS, G. (1992), 'Constrained restoration and the recovery of discontinuities', *IEEE Transactions on Image Processing*, Vol. 14, No. 3, March, pp. 367–383.

GERCHBERG, R.W. (1974), 'Super-resolution through error energy reduction', *Optica Acta*, Vol. 21, No. 9, pp. 709–720.

HAACKE, E.M., LIANG, Z.P. AND IZEN, S.H. (1989a), 'Superresolution reconstruction through object modeling and parameter estimation', Vol. 37, No. 4, April, pp. 592–595.

HAACKE, E.M., LIANG, Z.P. AND IZEN, S.H. (1989b), 'Constrained reconstruction: A superresolution, optimal signal-to-noise alternative to the Fourier transform in magnetic resonance imaging', *Medical Physics*, Vol. 16, No. 3, may/jun, pp. 388–397.

HART, H.R., BOTTOMLEY, P.A., EDELSTEIN, W.E., KARR, S.G., LEUE, W.M., MUELLER, O., REDINGTON, R.W., SCHENCK, J.F., SMITH, L.S. AND VATIS, D. (1983), 'Nuclear magnetic resonance imaging: contrast to noise ratio as a function of strength of magnetic field', *American Journal of Roentgenology*, Vol. 141, pp. 1195–1201.

HENKELMAN, R. AND J.BRONSKILL, M. (1987), 'Artifacts in magnetic resonance imaging', *Reviews of Magnetic Resonance in Medicine*, Vol. 2, No. 1, pp. 1–126.

HENNIG, J. (1999), 'K-space sampling strategies', *Europe Radiol.*, Vol. 9, pp. 1020–1031.

HIMMELBLAU, D.M. (1972), *Applied nonlinear programming*, McGraw-Hill Book Company.

HORNAK, J.P. (1996), 'The basics of MRI', http://www.cis.rit.edu/htbooks/mri/, Rochester Institute of Technology.

HOULT, D.I. AND LAUTERBUR, P.C. (1979), 'The sensitivity of the zeugmatographic experiment involving human samples', *Journal of Magnetic Resonance*, Vol. 34, pp. 425–433.

HU, X. (1994), 'On the keyhole technique', *Journal of Magnetic Resonance Imaging*, Vol. 4, p. 231.

HU, X. AND PARRISH, T. (1994), 'Reduction of the field of view for dynamic imaging', *Magnetic Resonance in Medicine*, Vol. 31, pp. 691–694.

HU, X., PARRISH, T. AND PATEL, M. (1994), 'Novel sampling schemes in magnetic resonance imaging', *Proceedings of IEEE International Conference on Image Processing VIII*, pp. 15–19.

HUBER, P.J. (1981), *Robust statistics*, John Wiley and Sons, Inc.

HUNT, B.R. (1973), 'The application of constrained least squares estimation to image restoration by digital computer', *IEEE Transactions on Computers*, Vol. C-22, No. 9, pp. 805–810.

HUNT, B.R. (1995), 'Super-resolution of images: algorithms, principles, performance', *International Journal of Imaging Systems and Technology*, Vol. 6, pp. 297–304.

HUTSON, V. AND PYM, J.S. (1980), *Application of functional analysis and operator theory*, Academic Press.

JAIN, A.K. (1989), *Fundamentals of digital image processing*, Prentice Hall.

JONES, R.A., HARALDSETH, O., MULLER, T.B., RINCK, P.A. AND OKSENDAL, A.N. (1993), 'K-space substitution: a novel dynamic imaging technique', *Magnetic Resonance in Medicine*, Vol. 29, pp. 830–834.

KANG, M.G. AND KATSAGGELOS, A.K. (1995), 'General choice of the regularization functional in regularized image restoration', *IEEE Transactions on Image Processing*, Vol. 4, No. 5, May, pp. 594–602.

KATSAGGELOS, A.K. (1989), 'Iterative image restoration algorithms', *Optical Engineering*, Vol. 28, No. 7, July, pp. 735–748.

KATSAGGELOS, A.K. (1991), *Digital image restoration*, Springer-Verlag.

KATSAGGELOS, A.K., BIEMOND, J., SCHAFER, R. AND MERSEREAU, R.M. (1991), 'A regularized iterative image restoration algorithm', *IEEE Transactions on Signal Processing*, Vol. 39, No. 4, pp. 914–929.

KIRSCH, A. (Ed.) (1996), *An introduction to the mathematical theory of inverse problems*, Springer-Verlag.

KWAN, R.K.S., EVANS, A.C. AND PIKE, G.B. (1996), 'An extensible MRI simulator for post-processing evaluation', *Visualization in Biomedical Computing (VBC'96), Lecture Notes in Computer Science*, Vol. 1131, pp. 135–140. (See also http://www.bic.mni.mcgill.ca/brainweb/).

KYRIAKOS, W.E., PANYCH, L.P., ZIENTARA, G.P. AND JOLESZ, F.A. (1997), 'Implementation of a reduced field-of-view method for dynamic MR imaging using navigator echoes', *Journal of Magnetic Resonance Imaging*, Vol. 7, pp. 376–381.

LI, S.Z. (1995), 'On discontinuity-adaptive smoothness priors in computer vision', *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 17, No. 6, pp. 576–586.

LI, S.Z. (1998), 'Close-form solution and parameter selection for convex minimization-based edge-preserving smoothing', *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 20, No. 9, pp. 916–932.

LIANG, Z.P. AND LAUTERBUR, P.C. (1994), 'An efficient method for dynamic magnetic resonance imaging', *IEEE Transactions on Medical Imaging*, Vol. 13, No. 4, December, pp. 677–686.

LIANG, Z.P. AND LAUTERBUR, P.C. (1996), 'Constrained imaging: overcoming the limitations of the Fourier series', *IEEE Engineering in Medicine and Biology Magazine*, Vol. 15, pp. 126–132.

LIANG, Z.P. AND LAUTERBUR, P.C. (1999), *Principles of magnetic resonance imaging, a signal procesing perspective*, IEEE Press.

LIANG, Z.P., HAACKE, E.M. AND THOMAS, C.W. (1988), 'High resolution spectral estimation through localized polynomial aproximation', *IEEE Proceddings On Acoustics, Speech, and Signal Processing 4th workshop Spectrum Est.*, August, pp. 402–407.

LIANG, Z.P., HAACKE, E.M. AND THOMAS, C.W. (1989), 'High-resolution inversion of finite Fourier transform data through a localised polynomial approximation', *Inverse Problems*, Vol. 5, pp. 831–847.

LORETTE, A., SHEKARFAROUSH, H. AND ZERUBIA, J. (1997), 'Super-resolution with adaptive regularization', *Proc ICIP*, October, pp. 169–172.

LUENBERGER, D.G. (1973), *Introduction to linear and nonlinear programming*, Addison-Wesley Publishing Company.

LUENBERGER, D.G. (1984), *Linear and nonlinear programming*, Addison-Wesley Publishing Company.

MARGOSIAN, P., SCHMITT, F. AND PURDY, D. (1986), 'Faster MR imaging: imaging with half the data', *Health Care Instrum*, Vol. 1, pp. 195–197.

MARTOS, B. (1975), *Nonlinear programming theory and methods*, American Elsevier Publising Co.

MCCORMICK, G.P. (1983), *Nonlinear programming: theory, algorithms, and applications*, John Wiley and Sons, Inc.

MILLER, K. (1970), 'Least squares methods for ill-posed problems with a prescribed bound', *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 1, No. 1, February, pp. 52–74.

MORRIS, P.G. (1986), *Nuclear magnetic resonance imaging in medicine and biology*, Clarendon Press.

NISHIMURA, D. (1994), *Principles of magnetic resonance imaging*, Stanford University.

PAPOULIS, A. (1975), 'A new algorithm in spectral analysis and band-limited extrapolation', *IEEE Transactions on Circuits and Systems*, Vol. 22, No. 9, September, pp. 735–742.

PHILIPS *Basic Principles of MR Imaging*, PHILIPS Medical Systems.

PRATT, W.K. (1978), *Digital image processing*, John Wiley and Sons, Inc.

PURCELL, E., TORREY, H. AND POUND, R. (1946), 'Resonance absorption by nuclear magnetic moments in a solid', *Physical Review*, Vol. 69, pp. 37–38.

RUDIN, L.I., OSHER, S. AND FATEMI, E. (1992), 'Nonlinear total variation noise removal algorithms', *Physica D*, Vol. 60, pp. 259–268.

S. LAWRENCE MARPLE, J. (1987), *Digital spectral analysis with applications*, Prentice Hall.

SAITO, N. (1990), 'Superresolution of noisy band-limited data by data adaptive regularization and its application to seismic trace inversion', *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1237–1240.

SEMENTILLI, P.J., N, M.S. AND HUNT, B.R. (1993), 'Poisson MAP super-resolution estimator with smoothness constraint', *SPIE*, Vol. 2032, pp. 2–13.

SEZAN, M.I. AND STARK, H. (1982), 'Image restoration by the method of convex projections: part 2 - applications and numerical results.', *IEEE Transactions on Medical Imaging*, Vol. MI-1, No. 2, October, pp. 95–101.

SEZAN, M.I. AND STARK, H. (1983), 'Image restoration by convex projections in the presence of noise.', *Applied Optics*, Vol. 22, No. 18, September, pp. 2781–2789.

SEZAN, M.I. AND STARK, H. (1984), 'Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion.', *IEEE Transactions on Medical Imaging*, Vol. MI-3, No. 2, June, pp. 91–98.

SHEWCHUCK, J.R. (1994), 'An introduction to the conjugate gradient method without the agonizing pain', http://www.cs.cmu.edu/afs/cs/project/quake/public/papers/painles-conjugate-gradient.abstract.

SHULMAN, D. AND HERVÉ, J.Y. (1989), 'Regularization of discontinous flow fields', *Proc. IEEE Workshop on Visual Motion*, March, pp. 81–86.

SLEPIAN, D. AND POLLACK, H.O. (1961), 'Prolate spheroidal wave functions, Fourier analysis and uncertainity-I', *Bell Syst. Tech. J.*, Vol. 40, pp. 43–63.

STARK, H. (Ed.) (1987), *Image recovery, theory and application*, Academic Press, INC.

STARK, H. (1990), 'Convex projections in image processing', *IEEE International Symposium on Circuits and Systems*, Vol. 3, March, pp. 2034–2036.

STEVENSON, R.L. AND DELP, E.J. (1990), 'Fitting curves with discontinuities', *IEEE Int. Workshop on Robust Computer Vision*, October, pp. 127–136.

STEVENSON, R.L., SCHMITZ, B.E. AND DELP, E.J. (1991), 'Surface reconstruction with dicontinuities', *SPIE Conference on Computer Vision and Graphics*, November.

STEVENSON, R.L., SCHMITZ, B.E. AND DELP, E.J. (1994), 'Discontinuity preserving regularization of inverse visual problems', *IEEE Transactions On Systems, Man and Cybernetics*, Vol. 24, No. 3, pp. 455–469.

STRATTON, J.A. (1941), *Electromagnetic theory*, McGraw-Hill Book Company.

STRONG, D.M. (1997), *Adaptive total variation minimizing image restoration*, PhD thesis, University of California, Los Angeles.

STRONG, D. AND CHAN, T. (1996a), 'Exact solutions to total variation regularization problems', *UCLA Computational and Applied Mathematics Reports*, pp. 1–33. (See also http://www.math.ucla.edu/applied/cam/index.html).

STRONG, D. AND CHAN, T. (1996b), 'Relation of regularization parameter and scale in total variation based image denoising', *UCLA Computational and Applied Mathematics Reports*, pp. 1–35. (See also http://www.math.ucla.edu/applied/cam/index.html).

SULLIVAN, B.J. AND LIU, B. (1984), 'On the use of singular value decomposition and decimation in discrete-time bandlimited signal extrapolation', *IEEE Proceedings On Acoustics, Speech, and Signal Processing*, Vol. 32, pp. 1201–1212.

THOMPSON, A.R., MORAN, J.M. AND JR., G.W.S. (1986), *Interferometry and synthesis in radio astronomy*, John Wiley and Sons.

TIKHONOV, A.N. AND ARSENIN, V.Y. (1977), *Solutions of ill-posed problems*, John Wiley and Sons, Inc.

VAN VAALS, J.J., BRUMMER, M.E., DIXON, W.T., TUITHOF, H.H., ENGELS, H., NELSON, R.C., GERETY, B.M., CHEZMAR, J.L. AND DEN BOER, J.A. (1993), 'Keyhole method for accelerating imaging of contrast agent uptake', *Journal of Magnetic Resonance Imaging*, Vol. 3, pp. 671–675.

VERVEER, P.J. AND JOVIN, T.M. (1997), 'Efficient superresolution restoration algorithms using maximum *a posteriori* estimations with application to fluorescence microscopy', *Journal of Optical Society of America A*, Vol. 14, No. 8, August, pp. 1696–1706.

VLAARDINGERBROEK, M.T. AND BOER, J.A.D. (1996), *Magnetic resonance imaging*, Springer-Verlag.

VOGEL, C.R. AND OMAN, M.E. (1996), 'Iterative methods for total variation denoising', *SIAM J. Sci. Comput.*, Vol. 17, March, pp. 227–238.

WALSH, D.O. AND NIELSEN-DELANEY, P.A. (1994), 'Direct method for superresolution', *Journal of Optical Society of America*, Vol. 11, No. 2, February, pp. 572–579.

WATKINS, D.S. (1991), *Fundamentals of matrix computations*, John Wiley and Sons, Inc.

WEBB, S. (1995), *The physics of medical imaging*, Institute of Physics Publishing.

WRIGHT, G.A. (1997), 'Magnetic resonance imaging', *IEEE Signal Processing Magazine*, January, pp. 56–66.

YEH, S. AND STARK, H. (1990), 'Iterative and one-step reconstruction from nonuniform samples by convex projections', *Journal of Optical Society of America A*, Vol. 7, No. 3, March, pp. 491–499.

YOULA, D.C. AND WEBB, H. (1982), 'Image restoration by the method of convex projections: part1 - theory', *IEEE Transactions on Medical Imaging*, Vol. MI-1, No. 2, October, pp. 81–94.